


# Limit setting using spacings in the presence of unknown backgrounds

Lolian Shtembari<sup>\*</sup> and Allen Caldwell<sup>Ⓜ</sup>

Max Planck Institute for Physics, Munich 80805, Germany

 (Received 20 March 2023; accepted 25 October 2023; published 4 December 2023)

Finding upper limits on the rate of events from a proposed process in the presence of unknown backgrounds is an often encountered problem in the search for rare processes. Methods based on unusually large “gaps,” or spacings, in the event distribution allow to set limits on the rate of the proposed signal distribution. In this paper, we present two novel spacings-based methods: the “sum of sorted spacings” and the “product of complementary spacings” as tests and compare these to existing tests on synthetic data as well as on a published dataset.

DOI: [10.1103/PhysRevD.108.123005](https://doi.org/10.1103/PhysRevD.108.123005)

## I. INTRODUCTION

Many experiments tasked with the discovery of theorized rare processes might find themselves in a situation where the collected data is insufficient to claim a positive detection. In such cases, the data is used to set an upper limit on the number of events resulting from the rare process under consideration, which in turn can be used to set an upper limit on physical quantities of the proposed model. An example would be the determinations of upper limits on the cross section of weakly interacting massive particles (WIMPs) recoiling off atoms in a detector, such as for the CRESST [1] or the CDMS [2] experiments. The experiments in question are often contaminated by a poorly understood background, in which case the signal strength limit must be set from properties of the observed event distribution without any background subtraction. Spacing statistics are one method to go beyond pure event counting in setting signal strength limits.

Since the expected shape of the event distribution produced by the targeted process is known, it is possible to estimate the number of events it accounts for, up to a desired confidence level, leveraging the difference between the observed event distribution and the expected one. Such an analysis is carried out using goodness-of-fit tests allowing for the assumption that the observed number of events collected in the analysis window is just a realization of a random variable following a Poisson distribution with unknown rate  $\mu$ . For a selected goodness-of-fit test, the goal

is to determine the event rate  $\mu$  coinciding with the desired confidence level.

In the following we give a brief review on how to use goodness-of-fit tests for an arbitrary event distribution assuming the targeted process and how to account for the random number of observed events in the definition of the  $p$  value.

We then discuss how to use spacings-based tests to provide upper limits: we begin with a quick review of the maximum gap and optimum interval methods [3] and then introduce two new tests based respectively on the sorted list of spacings and on the product of spacings.

## II. SETTING UPPER LIMITS WITH TEST STATISTICS

Several nonparametric goodness-of-fit tests exist to test a univariate distribution. Targeting vastly different univariate distributions is made possible by the probability integral transformation [4,5], which basically reduces the goodness of fit to a simple uniformity test.

### A. Probability integral transformation

Suppose we have  $n$  samples  $\{y_i\}$ , and want to quantitatively test the hypothesis of those samples being random variates of a known distribution  $f(y)$ , independent and identically distributed according to  $f(y)$ . Considering only continuous distributions  $f(y)$  with cumulative  $F(y)$ , it is possible to transform samples onto the unit interval  $[0, 1]$  via  $x_i = F(y_i)$ . This reduces the task at hand to test transformed samples  $\{x_i\}$  being distributed according to the standard uniform distribution  $\mathcal{U}(0, 1)$ .

### B. Poisson distribution and $p$ value

Given a dataset consisting of  $n$  uniformly distributed samples  $\{x_i\}$ , a test statistic  $T$ , and a function of the data  $t = g(\{x_i\})$ , the  $p$  value is directly calculable from  $F_T(t|n)$ ,

<sup>\*</sup>lolian@mpp.mpg.de

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.*

where  $F_T$  is the cumulative distribution function of the test  $T$  for exactly  $n$  events. Here, the p value treats the number of events in the analysis window as a fixed parameter. If we consider that the observed number of events is a random variable with an associated Poisson distribution, then it is possible to correct the definition of the p value by averaging over all possible numbers of events. Considering a Poisson distribution with rate  $\mu$ , and an observed test statistic value  $t_{\text{obs}} = g(\{x_i\})$ , the Poisson-averaged p value is calculated as

$$1 - p = F_{T,Pois}(t_{\text{obs}}|\mu) = \sum_{n=1}^{\infty} F_T(t_{\text{obs}}|n) \cdot \frac{\mu^n e^{-\mu}}{n!}, \quad (1)$$

where the sum starts at  $n = 1$  since the test statistic often is not defined for  $n = 0$ . In case of no observed events,  $n = 0$ , no improvement upon the simple Poisson statistic is possible, which becomes the extension of this approach in the limit of empty datasets.

### C. Setting upper limits

Given a test  $T$  and its distribution  $F_{T,Pois}$ , we showed above how to calculate the p value of a given dataset  $\{x_i\}$  comprised of  $n$  events, for a selected value of the event rate  $\mu$ . Instead of performing a goodness-of-fit test, assessing how well the event distribution fits that of a uniform distribution for a given  $\mu$ , we could determine which is the rate  $\mu$  that yields a desired p value, determining the event rate representative of the uniformly distributed subset of events up to a desired confidence level (C.L.).

As an example, for a given dataset  $\{x_i\}$  the upper limit on the event rate,  $\mu_U$ , at a C.L. is such that

$$1 - p = F_{T,Pois}(t_{\text{obs}}|\mu_U) = \text{C.L.} \quad (2)$$

## III. SPACING STATISTICS

Given a uniformly distributed and ordered set of  $n$  events  $\{x_i\}$  (where  $x_i < x_j \forall i < j$ ), in the interval  $[0, 1]$ , we can define the  $n + 1$  spacings  $s$  as  $s_{j,1} = x_j - x_{j-1}$ , with  $x_0 = 0$  and  $x_{n+1} = 1$ . These spacings are the gaps between consecutive events, but in general we could also consider higher order spacings: a spacing of order  $k$  would be  $s_{j,k} = x_j - x_{j-k}$ .

Based on these spacings it is possible to construct test statistics capable of setting much more competitive upper limits on the event rate than the simple counting (Poisson) test, since they not only consider the total number of data contained in the analysis window, but also their distribution, taking advantage of regions of relatively low event density in order to estimate the underlying uniform component of the event distribution.

### A. Maximum gap method

This maximum gap test has been proposed in the physics literature [3] (and earlier in the statistics literature [6]). It considers the largest spacing present in order to determine the upper limit on the event rate. The test statistic is defined as

$$T_{MG}(\{x_i\}) = s_{\text{max},1} = \max_i(s_{i,1}). \quad (3)$$

The distribution of  $s_{\text{max},1}$  for a given number of events is known [3,6]. Under the assumption of a uniform distribution of events, a large proposed event rate,  $\mu$ , will lead to a small probability to observe large values of  $s_{\text{max},1}$ . If the observed value of  $s_{\text{max},1}$  is indeed large relative to our expectations, this can be used to exclude values of  $\mu$  at a specified confidence level.

This definition of the test is particularly helpful since it is little affected by possible clustering of events in the unit interval, whose distribution deviates from the standard uniform one. This case arises when an experiment is afflicted by an unknown background and the probability integral transformation is performed with respect to the distribution of the signal under study: after the transformation the events deriving from a possible signal source will be uniformly distributed while events coming from unknown background sources will not follow a uniform distribution, producing regions in the unit interval with an overdensity of events and regions with large spacings.

The Poisson averaged cumulative distribution,  $F_{MG,Pois}$ , can be easily computed analytically and has been given by Yellin [3]:

$$F_{MG,Pois}(x|\mu) = \sum_{t=0}^m \frac{(tx - \mu)^t e^{-tx}}{t!} \left(1 + \frac{t}{\mu - tx}\right), \quad (4)$$

where  $m$  is the greatest integer  $\leq \mu/x$ . The C.L. upper limit on the event rate,  $\mu_{MG}$ , is such that

$$F_{MG,Pois}(s_{\text{max},1}|\mu_{MG}) = \text{C.L.} \quad (5)$$

### B. Optimum interval method

In addition to the maximum gap method, the optimum interval method has been proposed by Yellin [3] which instead of looking at the largest spacing, instead considers sums of ordered spacings: i.e., higher order spacings.

The maximum gap method compares the size of the largest first order spacing against the expectation of it containing no events for a given event rate  $\mu$ . Similarly, given higher order spacings, for example  $k = 2$ , we might find the largest second order spacing  $s_{\text{max},2}$  and compare its size to the expectation of it containing only one event given a proposed event rate  $\mu$ . Such an investigation can be performed for any order of spacings allowed by the data

( $k \leq n$ , where  $n$  is the number of events) and would result in  $n$  different Poisson-averaged p values, one for each order of spacing, for a given event rate  $\mu$ :

$$s_{\max,k} = \max_i(s_{i,k}) \quad (6)$$

$$1 - p_k = F_{O,k,Pois}(s_{\max,k}|\mu), \quad (7)$$

where  $F_{O,k}$  is the cumulative distribution of the  $s_{\max,k}$  for a given number of events. The analytic formula of  $F_{O,k}$  ( $k > 1$ ) for  $n$  events is not known, but Yellin calculated numerical approximations using a large Monte Carlo campaign as well as an approximate asymptotic distribution for large values of  $n$  [7].

In order to exclude the proposed event rate  $\mu$ , one might look at the smallest available p value:

$$p_{\min} = \min_k(p_k). \quad (8)$$

Since  $p_{\min}$  does not have a uniform distribution, one needs to know its cumulative distribution for a given event rate  $\mu$ ,  $F_{O,\min}$ , and the final p value is

$$p_{\text{fin}} = F_{O,\min}(p_{\min}|\mu). \quad (9)$$

The analytic formula of  $F_{O,\min}$  is not known, and a numerical approximation is derived using Monte Carlo simulations. Inverting the formula, the upper limit  $\mu_{OI}$  on the event rate up to a given C.L. is such that

$$F_{O,\min}(p_{\min}|\mu_{OI}) = 1 - \text{C.L.} \quad (10)$$

### C. Sum of sorted spacings method

The tests discussed so far were developed in order to be sensitive to the presence of any abnormally large gaps between the events. For relatively few events under analysis, there might be only one such abnormally large gap, in which case the maximum gap method might already provide the most stringent upper limit. If more than one such abnormally large gap is present, and if these gaps happen to be located near one another, then they can be integrated in one higher order spacing and the optimum interval method provides more competitive limits.

However, if the abnormally large gaps are interspaced by many small gaps, then the optimum interval method would lose sensitivity and not offer significant improvements compared to the maximum gap method: in these cases, the sum of ordered spacings approach saturates and is dominated by individual low order spacings.

Ideally, we would like to combine the abnormally large gaps individually, without the gaps placed between them. This approach would increase the sensitivity of the test and potentially allow setting more competitive upper limits. We now describe such an approach.

Given a set of  $n$  events  $\{x_i\}$  in the unit interval  $[0, 1]$ , one considers the set of first order spacings  $\{s_{i,1}\}$  and then proceeds to sort these from largest to smallest, constructing the set of  $n + 1$  elements  $\{g_i\}$  (where  $g_1 > g_2 > \dots > g_n > g_{n+1}$ ). Given such a set of sorted spacings,  $\{g_i\}$ , it is possible to consider higher order spacings summing over its elements. The  $k$ th ordered sum of sorted spacings,  $G_k$ , is just the sum of the  $k$  largest first order spacings:

$$G_k = \sum_{i=1}^k g_i. \quad (11)$$

In total there are up to  $n$  nontrivial  $G_k$  given  $n$  events, since the sum over all  $n + 1$  first order spacings is constrained to be equal to 1.

It is now possible to evaluate the p value of each order of sum of sorted spacings for a given event rate  $\mu$ :

$$1 - p_k = F_{S,k,Pois}(G_k|\mu), \quad (12)$$

where  $F_{S,k}$  is the cumulative distribution of the  $G_k$  for a given number of events. The analytic formula of  $F_{S,k}$  for  $n$  events is known: it was first derived by Mauldon [8]. We independently rederived it using an alternative approach reported in Appendix A.

As with the optimum interval method, it is possible to use the smallest p value,  $p_{\min}$ , as a test statistic in order to exclude a value of  $\mu$  that is too large:

$$p_{\min} = \min_k(p_k).$$

Since  $p_{\min}$  is not a valid p value any more, one needs to know its cumulative distribution for a given event rate  $\mu$ ,  $F_{S,\min}$ , and the final p value is calculated in this case as

$$p_{\text{fin}} = F_{S,\min}(p_{\min}|\mu). \quad (13)$$

The analytic formula of  $F_{S,\min}$  is not known, and a numerical approximation is found from Monte Carlo simulations. Inverting the formula, one needs to find the C.L. upper limit  $\mu_S$  such that

$$F_{S,\min}(p_{\min}|\mu_S) = 1 - \text{C.L.} \quad (14)$$

Although the analytic formula of  $F_{S,k}$  for  $n$  is available, it is not well behaved, since it relies on the iterative difference of extremely large numbers (as  $n$  increases), making it susceptible to catastrophic numerical cancellation when used on a computer. In order to avoid these problems, we computed the values of the function with high numerical precision on a suitable grid in order to construct a reliable monotonic cubic-spline interpolation [9] that can be used with default 64-bit floating-point arithmetic. Currently, these interpolations have been tabulated up to  $n = 700$

and allow the estimation of event rates up to  $\mu \lesssim 580$ , correspondingly limiting the number of events it is possible to analyze. The speed-up tables and the code used to compute the test statistic and the limits is available in the SPACINGSTATISTICS.JL [10] package for JULIA.

#### D. Product of complementary spacings method

Finally, we propose another test statistic that combines spacings between events regardless of their relative location. The stepping stone of our proposal is a test first proposed by Moran [11] which consists in the product of all the spacings between consecutive events:

$$T(\{x_i\}) = - \sum_{i=1}^{n+1} \log(s_{i,1}), \quad (15)$$

where  $s_{i,1}$  are the first order spacings and  $n$  is the number of events in the unit interval. This test was proposed as a goodness-of-fit test sensitive to clusters of data against the null hypothesis of a uniform distribution: the presence of small spacings will drive the whole product of spacings towards more extreme values.

In order to make this test sensitive to the presence of large spacings, we consider the complements of each first order spacing and take their product:

$$C(\{x_i\}) = - \sum_{i=1}^{n+1} \log(1 - s_{i,1}). \quad (16)$$

The distribution of this quantity for a fixed number of events  $n$ ,  $F_C(C|n)$ , is not known analytically, but we derived a numerical approximation based on Monte Carlo simulations and tabulated them for  $n \leq 10^4$ . Additionally, since the definition of the test is of the form  $\sum g(s_{i,1})$ , it is in general possible to derive its asymptotic distribution, as described by Darling [12] and LeCam [13], whose methods we used to find the limiting distribution of  $C$ . The asymptotic distribution of  $C$  as  $n \rightarrow \infty$  is

$$f_C(C|n \rightarrow \infty) = \mathcal{N}(n \cdot \mu_\infty, n \cdot \sigma_\infty), \quad (17)$$

where the formulas of  $\mu_\infty$  and  $\sigma_\infty$  are reported in Appendix B. Given this result we can use it to estimate the test statistic for any large value of  $n$  and effectively extend the applicability of this test and its limit calculation to large numbers of events.

The Poisson-averaged p value of this test for a given event rate  $\mu$  is simply

$$F_{C,Pois}(C|\mu) = \sum_{n=1}^{\infty} F_C(C|n) \cdot \frac{\mu^n e^{-\mu}}{n!}. \quad (18)$$

Inverting this formula, one finds the C.L. upper limit  $\mu_C$  such that

$$F_{C,Pois}(C|\mu_C) = \text{C.L.} \quad (19)$$

The tabulated distributions as well as the code used to compute the test statistic and the limits are available in the SPACINGSTATISTICS.JL [10] package for JULIA.

## IV. PERFORMANCE COMPARISON

The performance of our proposed methods was extensively studied using simulated examples. In our tests we introduced backgrounds of varying shapes and strengths.

In the following we compare our methods against the standard Poisson test and the optimum interval method, which is considered the state of the art for setting limits in experiments affected by an unknown background.

### A. Background-free experiment

We start considering the case in which no added background contaminates the experiment, in order to estimate the baseline of the different methods. For simplicity we chose a uniform distribution for the generation of the events, which coincides with the null hypothesis of all tests, and we vary the event rate used in the data generation. Figure 1 shows the median of the C.L. = 0.90 upper limits on the event rate set using different methods. In order to better discern differences between the efficiency of each method, we can look at the results normalized to the Poisson limit, as shown in Fig. 2. We notice that in this baseline scenario the Poisson test is the best of the bunch, as expected. Nevertheless, the results of the Poisson test do not drastically outperform any other.

### B. Exponential background-only experiment

Next we investigate the case in which a background is present in our simulations and the signal strength is negligible in comparison: this mimics a rare process search in which the signal might be absent. In our experiments we produce data directly in the cumulative space (hence the

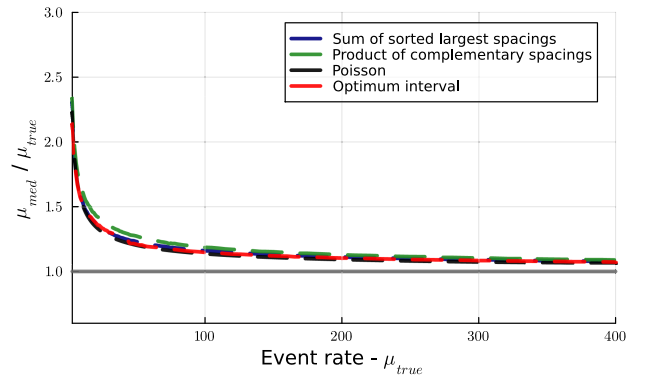


FIG. 1. Median C.L. = 0.90 upper limit normalized to the event rate used in the simulations; data generated according to a uniform distribution (background free).

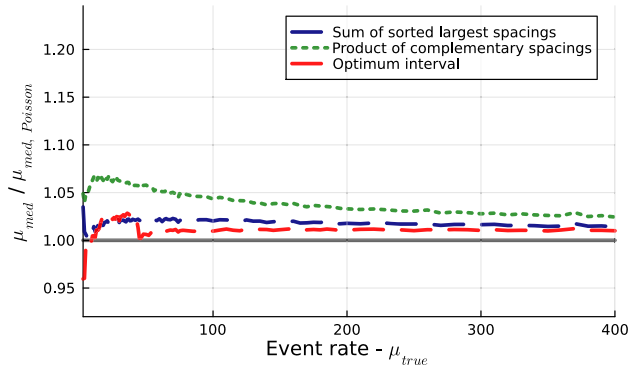


FIG. 2. Median C.L. = 0.90 upper limit normalized to the Poisson-test's result; data generated according to a purely uniform distribution (background free).

signal distribution is always assumed to be flat, i.e. the null hypothesis) and we start by considering an exponential background with rate 0.1 truncated on the unit interval  $[0, 1]$ . In dark matter search experiments it is often the case that the distribution of events, after transforming to the cumulative space, is peaked at one end of the analysis window with rapidly decaying tails. Figure 3 reports the median C.L. = 0.90 upper limits of the measured event rate normalized to the injected background event rate (bottom) as well as the ratio of the results of our methods against the optimum interval test (top). We have omitted the Poisson limit as it would simply scale with the total number of events and is not competitive. Analyzing the results, we notice that when dealing with relatively peaked event distributions all methods perform similarly, just like the background-free case. All methods are able to filter out most of the background contribution and reconstruct small overall event rates. For small injected background rates ( $\leq 100$ ), the nonlocal methods (sum of sorted spacings and

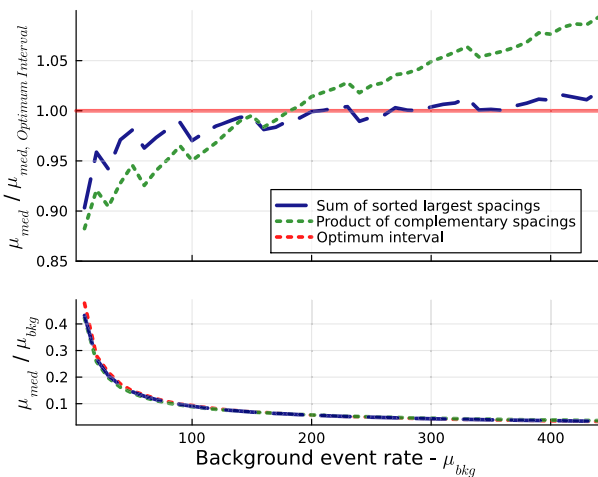


FIG. 3. Median C.L. = 0.90 upper limit normalized to the optimum interval result; data generated according to an exponential distribution of rate 0.1 (only background).

product of complementary spacings) are able to set up to 5%–10% more stringent limits. As the background rate increases ( $\geq 200$ ), the performance of the sum of the sorted spacings' test matches the optimum interval's one, while the product of the spacings' results are up to 5%–10% worse.

### C. Mixing background and signal

Now we investigate the case in which a detectable signal distribution is contaminated by background. In our simulations we will consider a background to signal ratio of 5, i.e. the rate of background events is 5 times larger than the rate of signal events. Since we operate directly in the cumulative space, we always assume a uniform signal distribution. The support of the background event distribution spans a quarter of the analysis window (meaning that the background events occupy a quarter of the unit interval) and we consider distributions of varying shape: specifically, the background distribution is a mixture of one or more uniform distributions whose total width sums up to 0.25.

Figure 4 shows different choices of background distributions in the left column, the resulting median C.L. = 0.90 limits for different methods, normalized to the optimum interval's result, in the central column.

If the background distribution is fully concentrated in one region, localized at either end of the analysis window, as shown in the first row of Fig. 4, this creates an uninterrupted low density region of the resulting event distribution. This is the best case scenario for the optimum interval method, as previously discussed. This expectation is reflected in the results, where the optimum interval method's results are up to 20% better than the other methods.

As we move the background distribution in the middle of the analysis window, or even split it up in two or more peaks, then we notice how our proposed tests are more sensitive, being able to set more competitive limits. For a bimodal background distribution it is possible to set limits 20% lower than the optimum interval method on average, while the gain rises up to 40% for a pentamodal background distribution. The performance of our proposed tests (middle column of Fig. 4) is due to their sensitivity to large regions of low event density, regardless of their number or location. With just two background peaks it is possible to be able to produce limits up to 20% or even 40% better than the optimum interval method as the low event density regions are further split.

Considering the case of a very faint or absent signal, we analyze the resulting limit if events were distributed only according to the background distribution. The results of these simulations are shown on the right column of Fig. 4, where we notice that, regardless of the shape of the background distributions, the 90% C.L. median results of the sum-of-sorted spacings and product-of-complementary spacings are always smaller than the optimum interval counterpart, with limit gains increasing up to a factor of 3 as the number of event-free regions increases.

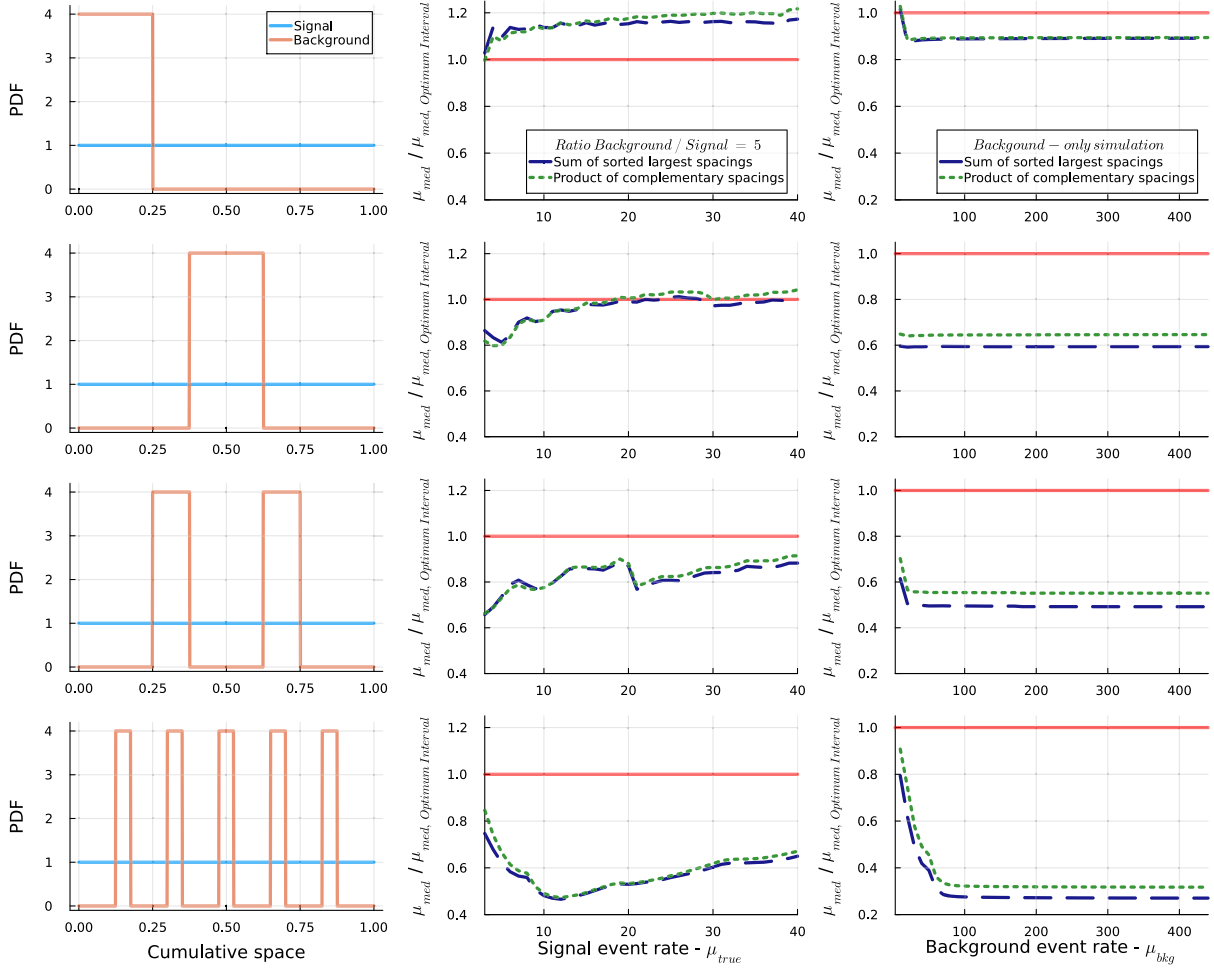


FIG. 4. Comparison of limit-setting methods depending on event distribution in the cumulative space: (left column) background and signal event distribution; (middle column) median C.L. = 0.90 upper limit normalized to the optimum interval’s result for simulated event distributions with a background and signal mixing of  $\mu_{\text{bkg}}/\mu_{\text{sig}} = 5$ ; (right column) median 90% C.L. upper limit normalized to the optimum interval’s result for purely background-like event distributions ( $\mu_{\text{bkg}}/\mu_{\text{sig}} = \infty$ ).

The case of multimodal background distributions, especially when it presents well-defined and relatively narrow peaks, is interesting since it is similar to experimental scenarios in which the event rate of a three or multibody decay is sought after: the expected spectrum of such a decay is relatively flat and could be contaminated by peaking background distributions which are representative of processes with a Standard Model counterpart. Our proposed methods would be well suited to tackle these problems since they are able to filter out the contributions coming from these “peaks” and estimate the underlying “flat” event rate, without introducing additional parameters in the analysis (biasing the result) to modify, limit or segment the region of interest in order to exclude peaking backgrounds.

#### D. Comparison to a likelihood-ratio test

As a further example, we compare the efficiency of the nonparametric tests discussed so far against a

likelihood-ratio (LR) test in the case of peaking backgrounds. We consider a Gaussian background with an associated event rate 10 times stronger than the signal’s event rate. The original distribution family of the background was fed into the LR method while the position, width, and the background to signal ratio were left as free parameters. Figure 5 shows the background distribution used and the median C.L. = 0.90 upper limits obtained at different signal event rates. Inspecting the results we notice that the most stringent limits are set by the LR approach, which is hardly surprising since partial information of the background was folded into the analysis. The limits set by the nonparametric spacings based tests, although more conservative, are still close enough to the LR ones, with our methods providing results no more than 20% larger than the LR for low event rates and up to 10% larger limits for high even rates. Furthermore, no assumption on the background shape is needed in the nonparametric tests.

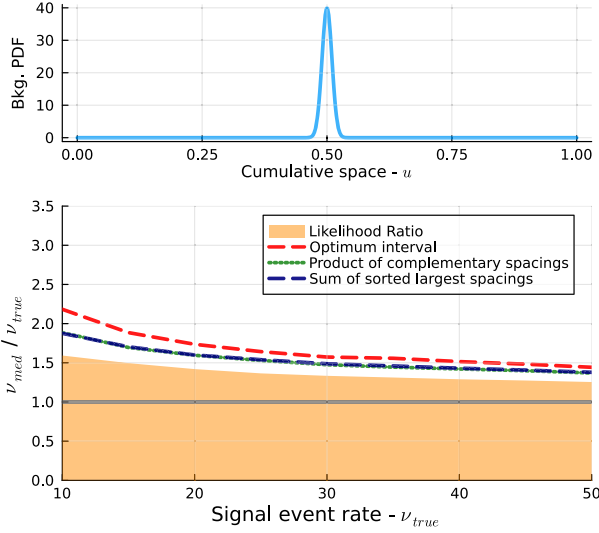


FIG. 5. Top: background distribution in the cumulative space: Gaussian distribution with  $\mu = 0.5$  and  $\sigma = 0.01$ ; Bottom: median C.L. = 0.90 upper limit normalized to the signal event rate.

## V. CRESST DATA EXAMPLE

We now replicate the analysis of the CRESST collaboration [1] to determine upper limits on the WIMP cross section. We use the most recent public dataset [14] released from the CRESST collaboration that is accompanied with information regarding the energy resolution and efficiencies of their setup for  $\text{CaWO}_4$  targets. The data we analyze are shown in Fig. 6 (top).

Given a specific WIMP mass and assuming a WIMP velocity distribution it is possible to calculate the differential rate  $\frac{dN}{dE}$  across an energy range of interest  $[E_{\min}, E_{\max}]$ . Denoting the integral of the differential rate over the whole energy range as  $\Lambda = \int_{E_{\min}}^{E_{\max}} \frac{dN}{dE} dE$ , for a given set of ordered events  $\{E_i\}$ , the probability integral transformation is simply

$$x_i = \frac{1}{\Lambda} \cdot \int_{E_{\min}}^{E_i} \frac{dN}{dE} dE \quad (20)$$

yielding a set of ordered events  $\{x_i\}$  in the unit interval  $[0, 1]$ .

Figure 6 (bottom) shows the distribution of events in the cumulative space after transforming using the signal distributions calculated at two different WIMP masses. Apart from small differences, the two datasets are very similar, showing an extremely peaked distribution close to 1 and an almost linear distribution of events in the rest of the unit interval.

Figure 7 shows the C.L. = 0.90 upper limits on the cross section we calculated using our methods as well as those computed with the optimum interval method, officially

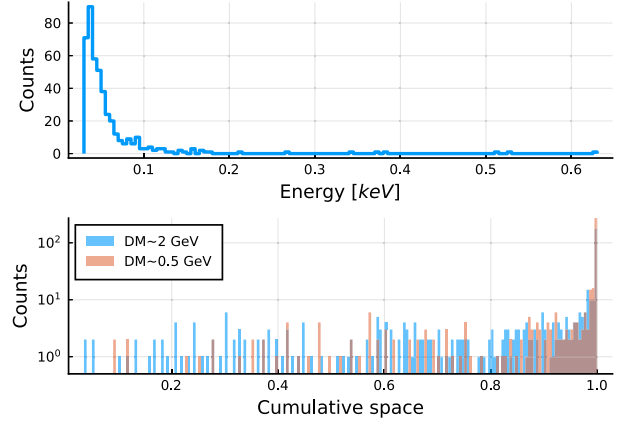


FIG. 6. Top: histogram of CRESST data [14] consisting of energy deposition from an interaction of a particle in the  $\text{CaWO}_4$  crystal. Bottom: histogram of data transformed using the signal distribution for two proposed WIMP masses, 0.5 and 2  $\text{GeV}/c^2$ .

used by the CRESST collaboration, which match the officially published limits.

Comparing the results of our calculations we notice there are no large deviations from one another. To better grasp the differences across results, we normalize the limits we obtain with our methods to the official ones (obtained with the optimum interval method), as shown in Fig. 8. Here, we notice that for the given data the product-of-complementary-spacings method yields 25% to 50% higher limits on average, whereas the sum-of-sorted-spacings method presents an oscillating behavior, being able to provide up to 30% lower limits for low WIMP masses and up to 40% higher limits for masses of the order of  $\sim 5 \text{ GeV}/c^2$ .

Finally, for WIMP masses  $\geq 20 \text{ GeV}/c^2$ , all methods saturate and yield the same result, reconstructing a signal event rate of  $\sim 2.3$  events, corresponding to the C.L. = 0.90 limit of the Poisson test for an empty analysis window.

This example based on a published dataset, as well as the results of our performance comparisons, shows that in general there is no “best” test statistic when it comes to

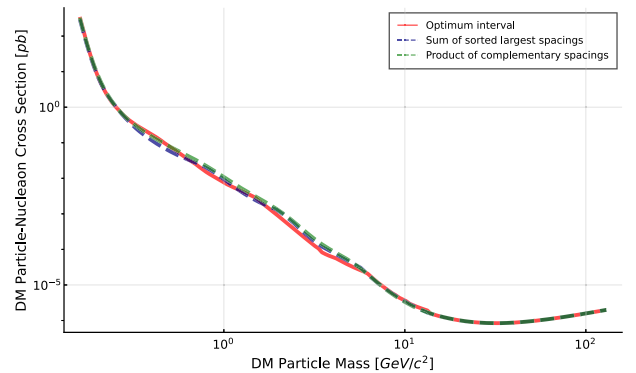


FIG. 7. C.L. = 0.90 upper limit on the WIMP-nucleon cross section as a function of the WIMP mass calculated with different tests.

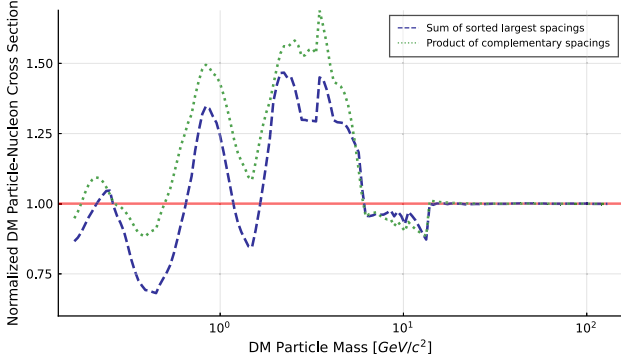


FIG. 8. C.L. = 0.90 upper limit on the WIMP-nucleon cross section normalized to the optimum interval's result as a function of the WIMP mass calculated with our proposed methods.

setting upper limits in the presence of unknown backgrounds, but the results are highly dependent on the actual distribution of events.

## VI. COMBINING UPPER LIMITS

When considering upper limits from experimental data, it is important to consider that often more than one experiment might produce a limit on the same quantity of interest, or even in a single experiment, one might have multiple detectors or multiple running periods. When a multitude of limits for the same quantity are available, one might be interested in estimating a combined limit. This issue has already been studied by Yellin and discussed in [15], where six different methods for combining limits have been presented. The methods presented in [15] refer to the optimum interval and maximum gap methods, but the behavior and distribution of the limits derived by the sum-of-sorted spacings and product-of-complementary spacings are the same, allowing the methods presented by Yellin to be applied to limits obtained with our tests. One additional improvement that can be taken advantage of by using the new tests presented here is that of analyzing the full range of data obtained in an experiment, without having to identify only the low-density regions and combining them via the serialization method: this is due to the feature of our new tests to automatically filter out high-density intervals and allows the total number of recorder events to be considered, without introducing any bias in the limit calculation.

## VII. CONCLUSIONS

In this paper we present two new methods to set upper limits in the presence of unknown backgrounds, namely the sum-of-sorted spacings and product-of-complementary spacings, and discuss their performance against the optimum interval method. The proposed tests leverage the presence of regions in the analysis window with low event density, regardless of their number or location relative to one another in order to estimate the underlying uniform

event distribution in the cumulative space. These features allow our tests to be viable alternatives for the analysis of rare process searches that aim to set competitive limits on their parameters of interest, especially when faced with peaked multimodal backgrounds.

## ACKNOWLEDGMENTS

We thank Dr. Heerak Banerjee, Dr. Oliver Schulz, and Dr. Nahuel Ferreiro Iachellini for the helpful discussions and comments that led to finalizing this work.

## APPENDIX A: DISTRIBUTION - SUM OF SORTED SPACINGS

Given  $m + 2$  ordered samples  $\{x_{(i)}\}$  in the interval  $[0, \mu]$ , where  $x_{(0)} = 0$  and  $x_{(m+1)} = \mu$ , we can define the  $m + 1$  spacings  $\{s_j\}$  as  $s_j = x_{(j)} - x_{(j-1)}$  and the ordered set of spacings  $\{s_{(j)}\}$  such that  $s_{(j)} < s_{(k)} \forall j < k$ .

The distribution of  $s_{(1)}$  is known:

$$P(s_{(1)} = x|m, \mu) = \frac{m(m+1)}{\mu} \left[ 1 - \frac{x(m+1)}{\mu} \right]^{m-1}. \quad (\text{A1})$$

The joint distribution of  $s_{(1)}$  and  $s_{(2)}$  can be written as follows:

$$P(s_{(1)} = x, s_{(2)} = y|m, \mu) = P(s_{(1)} = x|m, \mu) \cdot P(s_{(2)} = y|m, \mu, s_{(1)} = x). \quad (\text{A2})$$

Given the ordered set  $\{s_{(j)}\}$ , one can subtract  $s_{(1)}$  from all spacings and reduce the set to only  $m$  elements  $s_{(2)} - s_{(1)}, \dots, s_{(m+1)} - s_{(1)}$  which sum up to  $\mu - (m+1) \cdot s_{(1)}$ . Given this reduced set,  $s_{(2)} - s_{(1)}$  is the new smallest elements, whose distributions is given by Eq. (A1). Thus, we can rewrite  $P(s_{(1)} = x, s_{(2)} = y)$  as follows:

$$\begin{aligned} P(s_{(1)} = x, s_{(2)} = y|m, \mu) &= P(s_{(1)} = x|m, \mu) \cdot P(s_{(1)} \\ &= y - x|m - 1, \mu - (m+1)x). \end{aligned} \quad (\text{A3})$$

Given  $P(s_{(1)} = x, s_{(2)} = y)$ , we can integrate over its support to marginalize  $P(s_{(1)} + s_{(2)} = z)$ .

We define the sum of the  $k$  smallest spacings as  $S_k$ :  $S_k = \sum_{j=1}^k s_{(j)}$ . The joint distribution of  $P(s_{(1)} = x, g_k = y)$  can be expressed as

$$\begin{aligned} P(s_{(1)} = x, S_k = y|m, \mu) &= P(s_{(1)} = x|m, \mu) \cdot P(S_k = y|m, \mu, s_{(1)} = x) \\ &= P(s_{(1)} = x|m, \mu) \cdot P(S_{k-1} = y - kx|m - 1, \mu - (m+1)x). \end{aligned} \quad (\text{A4})$$

Assuming we know  $P(s_{k-1} = x)$  we can marginalize over  $s_{(1)}$  in order to obtain the distribution of  $g_k$ . This shows



a recursion between the distributions of  $S_{k-1}$  and  $S_k$ , which allows us to formulate a hypothesis on the distribution of  $S_k$  which can be proven by induction. The distribution of  $S_k$  we find is

$$P(S_k = x|n, 1) = A(k, n) \cdot \sum_{j=1}^k a(j, k) \left[ 1 - \left( \frac{n+2-j}{k+1-j} \right) x \right]^{n-1} h(x, j, k, n) \quad (\text{A5})$$

with

$$A(k, n) = \frac{n(n+1)!}{(n+1-k)^{k-1}(n+1-k)!} \quad (\text{A6})$$

$$a(j, k) = \frac{(-1)^{j-1}(k+1-j)^{k-2}}{(k-j)!(j-1)!} \quad (\text{A7})$$

$$h(x, j, k, n) = H(x) - H\left(x - \frac{k+1-j}{n+2-j}\right), \quad (\text{A8})$$

where  $H(x)$  is the Heaviside step function.

Finally, since all spacings  $s_{(j)}$  sum up to one in the unit interval, the sum of the  $m+1-k$  smallest spacings is the complement of the sum of  $k$  largest spacings  $G_k = \sum_{j=m+2-k}^{m+1} s_{(j)}$ . Thus, the distribution of  $g_k$  is simply

$$P(G_k = x|m, 1) = P(S_{m+1-k} = 1-x|m, 1). \quad (\text{A9})$$

## APPENDIX B: ASYMPTOTIC DISTRIBUTION - PRODUCT OF COMPLEMENTARY SPACINGS

Using LeCam's theorem [13], we are able to derive the asymptotic distribution of the “product of complementary spacings” test statistic. As the number of samples  $n \rightarrow \infty$ , the asymptotic distribution is

$$f_C(C|n \rightarrow \infty) = \mathcal{N}(n \cdot \mu_\infty, n \cdot \sigma_\infty), \quad (\text{B1})$$

where the parameters are given by

$$\mu_\infty(x) = e^{-x}[E_1(-x) - 2i\pi] \quad (\text{B2})$$

$$\sigma_\infty^2(x) = e^{-x}[2A(x) - 4i\pi B(x) - 2xe^{-x}E_1(-x)] - 1 - (x^2 + 1)e^{-2x}C(x), \quad (\text{B3})$$

where

$$A(x) = xF\left(\begin{matrix} [1, 1, 1] \\ [2, 2, 2] \end{matrix}; x\right) + \frac{\ln^2(-x)}{2} + \gamma \ln(-x) + \frac{\pi^2}{12} + \frac{\gamma^2}{2} \quad (\text{B4})$$

$$B(x) = \gamma + \ln(x) + x \quad (\text{B5})$$

$$C(x) = [e^{-x}E_1(-x)]^2 + 4i\pi e^{-x}E_1(-x) - 4\pi \quad (\text{B6})$$

with  $F$  being the hypergeometric function and  $\gamma$  the Euler-Mascheroni constant.

- 
- [1] M. Bravin *et al.* (CRESST Collaboration), *Astropart. Phys.* **12**, 107 (1999).  
[2] D. Abrams *et al.* (CDMS Collaboration), *Phys. Rev. D* **66**, 122003 (2002).  
[3] S. Yellin, *Phys. Rev. D* **66**, 032005 (2002).  
[4] K. Pearson, *Biometrika* **1**, 390 (1933).  
[5] K. Pearson, *Biometrika* **25**, 379 (1933).  
[6] R. A. Fisher, *Proc. R. Soc. A* **125**, 54 (1929).  
[7] S. Yellin, [arXiv:0709.2701](https://arxiv.org/abs/0709.2701).  
[8] J. G. Mauldon, *Math. Proc. Cambridge Philos. Soc.* **47**, 331 (1951).  
[9] F. N. Fritsch and R. E. Carlson, *SIAM J. Numer. Anal.* **17**, 238 (1980).  
[10] L. Shtembari, SPACINGSTATISTICS.JL (2023), <https://github.com/bat/SpacingStatistics.jl/tree/dev>.  
[11] P. A. P. Moran, *J. R. Stat. Soc. Ser. B* **13**, 147 (2018).  
[12] D. A. Darling, *Ann. Math. Stat.* **24**, 239 (1953).  
[13] L. L. Cam, *Publ. Inst. Statist. Univ. Paris* **7**, 7 (1958).  
[14] A. H. Abdelhameed *et al.* (CRESST Collaboration), *Phys. Rev. D* **100**, 102002 (2019).  
[15] S. Yellin, [arXiv:1105.2928](https://arxiv.org/abs/1105.2928).