Modifications of the Syrjala Test for Testing Spatial Distribution Differences Between Two Populations

Eric McKinney * Jürgen Symanzik †

Abstract
Determining whether two spatial distributions are statistically equivalent is the goal of the Syrjala test. When using continuous bivariate data, we show that the original Syrjala test produces different results depending on the data aggregation steps. In this article, we propose modifications to the previous version of the Syrjala test and make comparisons via simulations and an application. Simulation results indicate greater power and a more appropriate type one error rate for our modified Syrjala test. Furthermore, our new approach can be used for environmental data (for which the Syrjala test was originally developed), but also for data that originates from an eye-tracking study conducted at Utah State University.

Key Words: Continuous Bivariate Data; Permutation Test; Environmental Data; Eye-Tracking Data

1. Introduction

When handling two samples of continuous bivariate data, the question arises as to whether the sampled data come from the same population. Well known tests for determining whether two samples come from the same population in the univariate case include the Kolmogorov-Smirnov test (Kolmogorov, 1933) and the Cramér-von Mises test (Cramér, 1928; von Mises, 1928). Syrjala (1996) proposed a generalization of the Cramér-von Mises test to the bivariate case. The Syrjala test has been applied in many cases in the literature including tests of differences in the spatial distributions of adult vs. juvenile Pacific cod off the coast of Alaska (Syrjala, 1996), tests of differences in the distribution of the same bird species over three consecutive years in Central Spain (Benayas et al., 2010), and tests of differences in the distribution of two different respiratory infections affecting turtles in the Mojave Desert (Berry et al., 2015). Chiu and Liu (2009) summarized several other proposed generalizations of the Cramér-von Mises test.

In some cases, due to restrictions on the sampling locations being identical within the Syrjala test, preliminary data binning has been carried out (Chetverikov et al., 2018; McAdam et al., 2012). However, we show that depending on the data aggregation steps (such as binning) results of the Syrjala test can be contradictory. We propose modifications to the Syrjala test which eliminate the restriction for the sampling locations to be identical and the need to bin the data. A simulation study suggests that our modified version of the Syrjala test is in general more powerful and more appropriately sensitive.

Furthermore, an application of both the Syrjala test and our modified Syrjala test to a study involving eye-tracking and posture perception of individuals at Utah State University (USU) (Symanzik et al., 2017, 2018) showed stable results in our modified Syrjala test as compared to differing results from the original Syrjala test depending on the data binning technique employed.

*Department of Mathematics and Statistics, Utah State University, Logan, UT 84322–3900, USA. E-mail: ericmckinney77@gmail.com

†Department of Mathematics and Statistics, Utah State University, Logan, UT 84322–3900, USA. E-mail: symanzik@math.usu.edu

This article is structured as follows: In Section 2, we provide a background of the Cramér-von Mises test for differences in two sample empirical distributions. In Section 3, we introduce Syrjala's test and outline its original use for spatial data. In Section 4, modifications to the Syrjala test are proposed and discussed. Section 5 outlines the setup and results of a simulation study comparing the Syrjala test with our proposed modified Syrjala test. Section 6 contrasts the results of the two tests applied to eye-tracking data obtained from a posture study at USU. In Section 7, we provide conclusions and make note of additional elements for future study. All of our visualizations and analyses are conducted with the `R` statistical computing platform (R Core Team, 2019).

2. Tests for Identical Bivariate Distributions

Harald Cramér and Richard Edler von Mises proposed a test for determining whether a univariate sample comes from a theoretical distribution (Cramér, 1928; von Mises, 1928). Anderson (1962) generalized the Cramér-von Mises test to the two sample setting where the test determines whether two samples come from the same distribution (or whether there is some unspecified difference between the two samples). Specifically, let $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ and $X_{2,1}, X_{2,2}, \ldots, X_{2,m}$ be two independent random samples with unknown distribution functions $F_1(x)$ and $F_2(x)$ and empirical distribution functions $S_1(x)$ and $S_2(x)$, respectively. Then the hypotheses under consideration are as follows:

H_0: $F_1(x) = F_2(x) \quad \forall x$
H_a: $F_1(x) \neq F_2(x)$ for at least one value of x.

For comparison to our modified Syrjala test (in Section 4), Conover (1998) showed that Anderson's extension to the Cramér-von Mises statistic can be written as

$$T = \frac{mn}{(m+n)^2} \left\{ \sum_{i=1}^{n} \left[S_1(x_{1,i}) - S_2(x_{1,i})\right]^2 + \sum_{j=1}^{m} \left[S_1(x_{2,j}) - S_2(x_{2,j})\right]^2 \right\}. \quad (1)$$

Since the Cramér-von Mises test is a permutation test (also called a randomization test, re-randomization test, or an exact test) (Berry et al., 2011), the test statistic T_i is recalculated $N = \frac{(n+m)!}{n!m!}$ times, where $i = 1, \ldots, N$. Specifically, N is the total number of permutations of the sample labeling subscripts within $X_{1,1}, X_{1,2}, \ldots, X_{1,n}$ and $X_{2,1}, X_{2,2}, \ldots, X_{2,m}$, where each permutation will relabel n points as sample one and m points as sample two. The p-value is calculated as the total proportion of test statistics T_i which are greater than or equal to the statistic T computed from the non-permuted data, i.e.,

$$p - value = \frac{\sum_{i=1}^{N} I(T_i \geq T)}{N}, \quad (2)$$

where I is the indicator function.

3. The Syrjala Test

Stephen E. Syrjala proposed a test for differences in the spatial distribution of two groups as a generalization of the univariate Cramér-von Mises test to the bivariate setting (Syrjala, 1996). The test is designed for determining whether there is a difference between the locations of two populations at a single point in time, or whether there is a difference between the locations of the same population at two points in time. However, the test requires that

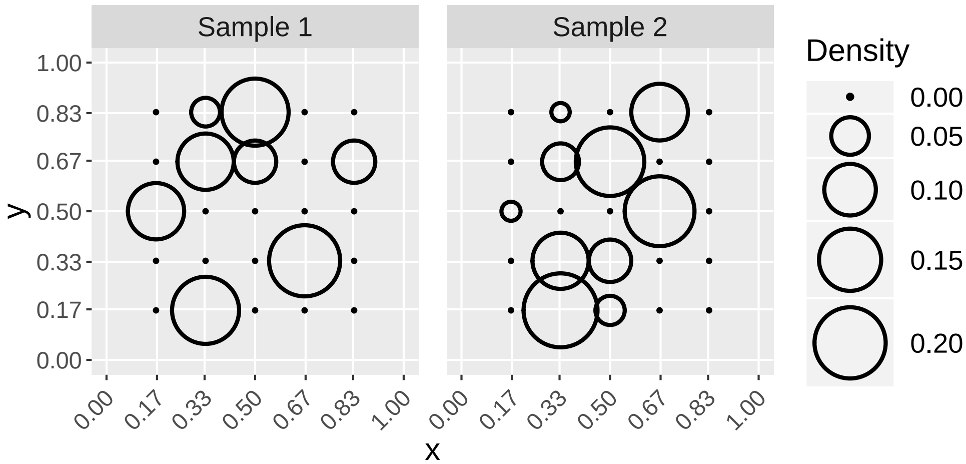

Figure 1: An example of identical sampling locations with differing location densities from two spatial distributions inspired by Benayas et al. (2010), page 312. Larger radii of the circles represent greater densities at each of the sampling locations.

the two samples both occur at an identical set of predefined locations. Furthermore, “The random variable in this case is the observed density at the sampling location, not the location itself.” (Syrjala, 1996). Hence, the hypotheses under consideration for the Syrjala test can be stated as

H_0: The normalized distributions of the populations are equal across the study area.
H_a: There is some unspecified difference in the normalized population distributions.

Borrowing from Syrjala (1996), let $d_t(x_k, y_k)$ denote the density of observations for group t; $t = 1, 2$, at sample locations (x_k, y_k); $k = 1, \ldots, K$, relative to their position on a bounding rectangle $\mathcal{A}$, where K is the total number of sampling locations. Then, $D_t = \sum_{k=1}^{K} d_t(x_k, y_k)$ is the sum of all densities across $\mathcal{A}$, and $\gamma_t(x_k, y_k) = \frac{d_t(x_k,y_k)}{D_t}$ are the normalized densities. With these, we can construct $\Gamma_t(x, y) = \sum_{x_k \leq x, y_k \leq y} \gamma_t(x_k, y_k)$ which is analogous to the bivariate empirical distribution function. Figure 1 shows a visualization of densities of observations at each of the identical sampling locations for two separate samples.

Similar to the Cramér-von Mises permutation test, the Syrjala first calculates a statistic (Ψ) on the non-permuted samples.

$$\Psi = \sum_{k=1}^{K} \left[\Gamma_1(x_k, y_k) - \Gamma_2(x_k, y_k)\right]^2 . \quad (3)$$

Because this construction uses points near the origin more often than points in the center, a common adjustment to Syrjala’s test is

$$\widetilde{\Psi} = \frac{1}{4} \sum_{c=1}^{4} \Psi_c, \text{where } \Psi_c = \sum_{k=1}^{K} \left[\Gamma_1(x_{c,k}, y_{c,k}) - \Gamma_2(x_{c,k}, y_{c,k})\right]^2 , \quad (4)$$

and $(x_{c,k}, y_{c,k})$ are positive coordinates defined relative to each of the four corners of $\mathcal{A}$ (Syrjala, 1996).

Next, permutations of the data are made by choosing to swap or leave $\gamma_1(x_k, y_k)$ and $\gamma_2(x_k, y_k)$ at each (x_k, y_k) locations. This results in $M = 2^K$ possible permutations of the data. Test statistics $\widetilde{\Psi}_j; j = 1, \ldots, M$, are calculated for each of the permutations of the data. The $\widetilde{\Psi}_j$ calculations are identical to that of $\widetilde{\Psi}$ (including the four rotations) except that they are computed from the permuted data. The p-value is calculated as the proportion of permuted statistics $\widetilde{\Psi}_j$ which are greater than or equal to the original statistic $\widetilde{\Psi}$ that is computed from the non-permuted data, i.e.,

$$p - value = \frac{\sum_{j=1}^{M} I(\widetilde{\Psi}_j \geq \widetilde{\Psi})}{M}. \tag{5}$$

It should be noted that in practice only a subset of $M' \ll M$ (typically $M' \approx 1000$) possible permutations are used to calculate the level of significance due to computational limitations.

4. Modifications of the Syrjala Test

Although binning of data to establish the common sampling locations required by the Syrjala test has been used in the literature (Chetverikov et al., 2018; McAdam et al., 2012), we show that the results of the Syrjala test depend on the binning technique (see Section 5.1 for more details). Hence, we propose modifications to the Syrjala test which eliminate the identical sampling locations restriction.

Extending our previous notation, let $(X_{1,1}, Y_{1,1}), (X_{1,2}, Y_{1,2}), \ldots, (X_{1,n}, Y_{1,n})$ and $(X_{2,1}, Y_{2,1}), (X_{2,2}, Y_{2,2}), \ldots, (X_{2,m}, Y_{2,m})$ be two independent random samples with unknown distribution functions $F_1(x, y)$ and $F_2(x, y)$ and bivariate empirical cumulative distribution functions (ecdf) $\Gamma_1^*(x, y)$ and $\Gamma_2^*(x, y)$, respectively. Then the hypotheses under consideration are as follows:

H_0: $F_1(x, y) = F_2(x, y) \quad \forall(x, y)$
H_a: $F_1(x, y) \neq F_2(x, y)$ for at least one coordinate pair (x, y).

In contrast to the Syrjala test, $\Gamma_1^*(x, y)$ and $\Gamma_2^*(x, y)$ in this test evaluate at each sampling location within their respective samples instead of at identical sampling locations from the two samples. Furthermore, the data will be rotated R times (instead of four times). Hence, the test statistic can be written as

$$\begin{aligned} \Psi^* = \sum_{r=1}^{R} \frac{1}{R} \Bigg\{ & \frac{n}{(m+n)} \sum_{i=1}^{n} \left[\Gamma_{1,r}^*(x_{1,i}, y_{1,i}) - \Gamma_{2,r}^*(x_{1,i}, y_{1,i})\right]^2 \\ & + \frac{m}{(m+n)} \sum_{j=1}^{m} \left[\Gamma_{1,r}^*(x_{2,j}, y_{2,j}) - \Gamma_{2,r}^*(x_{2,j}, y_{2,j})\right]^2 \Bigg\}, \end{aligned} \tag{6}$$

where R is a discrete number of rotations within 360°.

Our modified test statistic (Ψ^*) first computes the squared difference between the bivariate ecdfs evaluated at all of the data from the first sample. This sum of squared differences is weighted depending on the amount of data contributed from the first sample. This process is repeated for the second sample, and the two weighted sums of squared differences are then combined. Next, the data are rotated $360/R$ degrees, and another weighted average of the sums of squared differences of the ecdf values is computed. This computation is repeated for a total of R rotations, using $\Gamma_{1,r}^*$ and $\Gamma_{2,r}^*$ for each of the r^{th} rotations, with each rotation being weighted by $1/R$. This is the computation of the test statistic on the original data.

As a permutation test, the test statistic $\Psi_l^*; l = 1, \ldots, N$, is recalculated $N = \frac{(n+m)!}{n!m!}$ times where n and m are the respective sample sizes, and N is the total number of permutations of the sample labeling subscripts within $(X_{1,1}, Y_{1,1}), (X_{1,2}, Y_{1,2}), \ldots, (X_{1,n}, Y_{1,n})$ and $(X_{2,1}, Y_{2,1}), (X_{2,2}, Y_{2,2}), \ldots, (X_{2,m}, Y_{2,m})$. The p-value is calculated as the total proportion of test statistics Ψ_l^* which are greater than or equal to the statistic Ψ^* computed from the non-permuted data, i.e.,

$$p - value = \frac{\sum_{l=1}^{N} I(\Psi_l^* \geq \Psi^*)}{N}. \tag{7}$$

A visualization of the calculation of our modified Syrjala test can be seen in Figure 2. The top left graph in Figure 2 highlights three points from the two samples. The highlighted vertical bars seen between the two ecdfs in the bottom left graph represent the differences between the ecdfs evaluated at the respective highlighted points. The remaining two columns in Figure 2 suggest similarly made calculations (on the same highlighted points), but for rotated versions of the data. In this case, the data are being rotated every $40°$ for a total of $R = 360/40 = 9$ rotations. However, only the first two rotations are shown in Figure 2.

It should be noted that the bottom row of graphs in Figure 2 displays only the marginal ecdfs for each sample (and not the bivariate ecdfs). However, the difference between overlapping bivariate ecdfs is difficult to visualize. Hence, the marginal ecdfs are shown for visualization purposes only. Figure 3 compares the two bivariate ecdfs for the same data used in Figure 2.

5. Simulation

5.1 Simulation Design

A simulation study was carried out to compare the performances of the Syrjala test with our proposed modified Syrjala test introduced in Section 4. To assess the tests when the null hypothesis is true, two realizations of independent, uniformly distributed, or completely spatially random (CSR), data were simulated on $[0, 1] \times [0, 1]$ square regions. To assess the tests when the null hypothesis is false, four separate comparisons were made, each of which was compared to CSR. The four departures from CSR (also simulated on the $[0, 1] \times [0, 1]$ square) were constructed using the following intensity functions for the heterogeneous Poisson process where the values a_1, a_2, a_3, and a_4 are height parameters.

$$f_1(x, y) = a_1 \cdot \exp\Big\{ -20 \cdot \big[(x - 0.5)^2 + (y - 0.5)^2\big]\Big\} \qquad \text{(Center)}$$

$$f_2(x, y) = a_2 \cdot \Big(1 - \exp\Big\{ -80 \cdot \big[(x - 0.5)^4 + (y - 0.5)^4\big]\Big\}\Big) \qquad \text{(Repel)}$$

$$f_3(x, y) = a_3 \cdot \exp\Big\{ -5 \cdot \big[(x - 1)^2 + (y - 1)^2\big]\Big\} \qquad \text{(Corner)}$$

$$f_4(x, y) = a_4 \cdot \exp\Big\{ -5 \cdot (x - 1)^2\Big\} \qquad \text{(Right)}$$

Let μ be the average number of points within the unit square for the heterogeneous Poisson process. For reproducibility, Table 1 shows the values for the height parameters

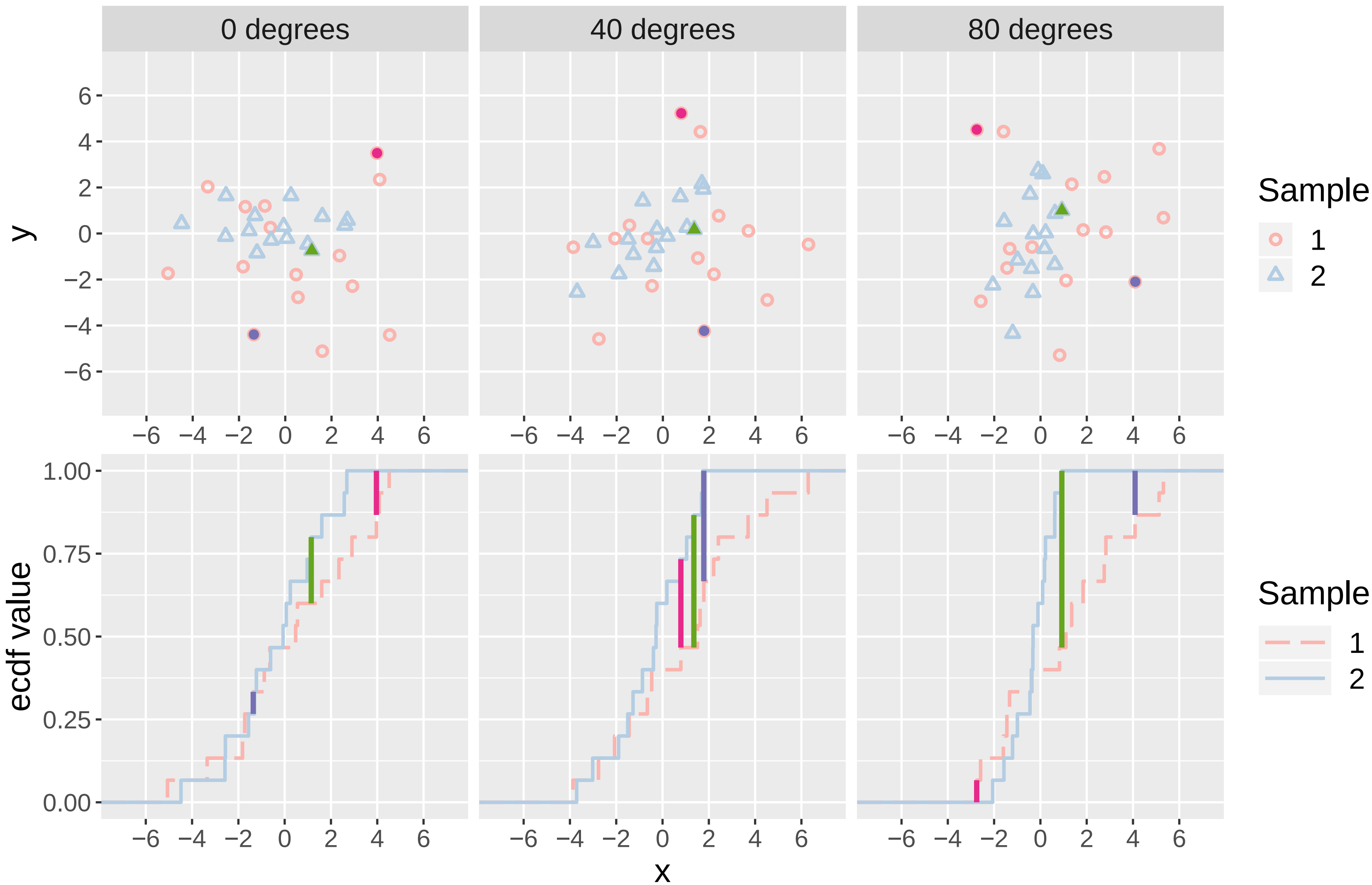

Figure 2: A visualization of calculations within our modified Syrjala test statistic. The same three demonstrative colored points (two from sample one, and one from sample two) are highlighted in the scatter plots (top row) across three different rotations of the data. The bottom row of graphs highlights three differences (vertical colored bars) between the ecdfs. Each ecdf difference (below) corresponds to a highlighted scatter plot point (above). While only three points and differences are highlighted, the calculation involves squared differences between ecdfs across all of the points from both samples. The bottom row shows differences between the marginal (and not bivariate) ecdfs. This is due to the difficult nature of visualizing differences in overlapping bivariate ecdfs. Hence, the marginal ecdfs are displayed for visualization purposes only. A comparison of the bivariate ecdfs is shown in Figure 3.

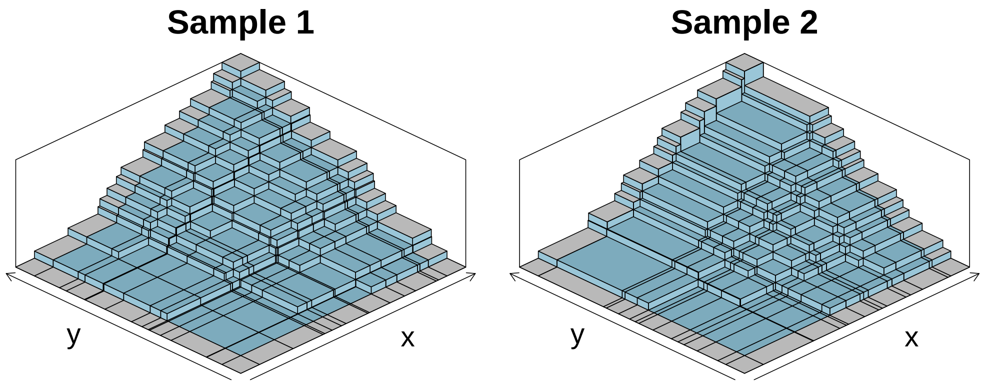

Figure 3: A visualization of the two bivariate ecdfs for the non-rotated samples shown in Figure 2.

Table 1: A table of the height parameter values (a_1, a_2, a_3, and a_4) which achieve a desired average number of points within the unit square (μ) for each respective intensity function.

μ	a_1	a_2	a_3	a_4
50	319	79	319	126
100	639	158	639	253
250	1597	395	1597	632
500	3193	790	3193	1264

a_1, a_2, a_3, and a_4 that achieve a specified intensity μ for each departure from CSR. The coefficients within the exponents of each of the intensity functions were chosen to ensure a sufficient departure from CSR was simulated. These coefficients also guarantee at least 97% of the area under each intensity function lie within the unit square. For each of the five comparisons (CSR compared with CSR, Center, Repel, Corner, Right), a CSR realization of 500 points was compared to four different sample sizes for each of the comparison distributions, specifically, sample sizes of 50, 100, 250, and 500. The results from the simulation study are summarized and displayed as a grid of line graphs in Figure 4. Visualizations of the CSR and heterogeneous Poisson process realizations (with $\mu = 500$ points) using each one of the intensity functions (referred to as CSR, Center, Repel, Corner, and Right, respectively) can be seen as a column of graphs in the far right of Figure 4.

Before applying the Syrjala test, two different types of binning were applied to the data, namely regular and random binning. Regular binning consists of dividing the sample region into a grid of equally sized rectangles. The density of all sample points within each rectangle was reported at the center of the respective rectangles. Random binning consists of randomly assigning binning points (using a simple sequential inhibition process) across the sample region, and assigning each sample point to the closest random binning point (using Euclidean distance). Within each of these binning approaches, three levels of granularity were used. Regular binning consisted of dividing the unit square into 5×5, 10×10, and 20×20 rectangular grids. Random binning involved randomly assigning 25, 100, and 400 random binning points across the sample region.

For our modified Syrjala test, by default no binning of the data was required. However, a subjective number of rotations to the data must be set. Hence, the Syrjala test was applied

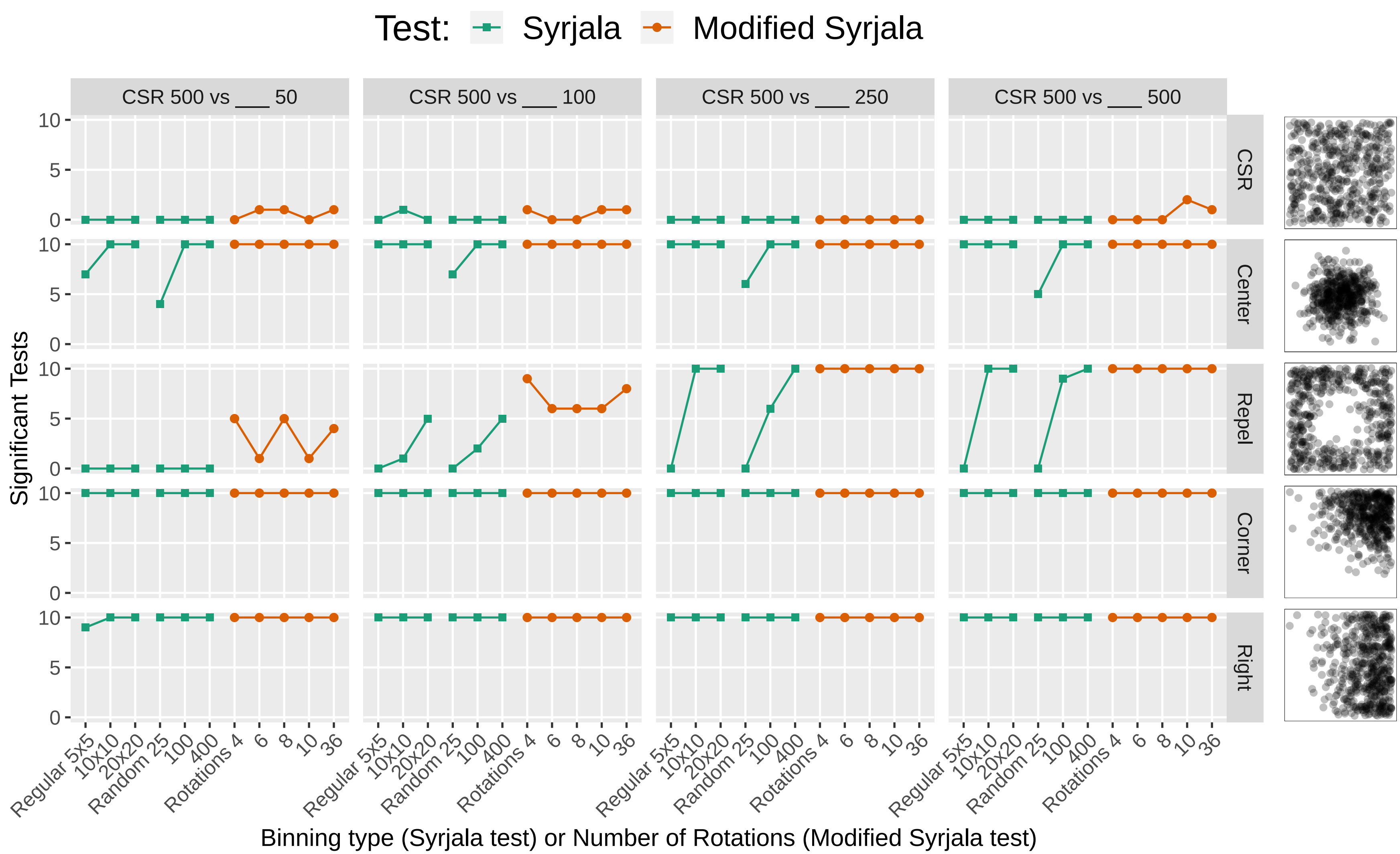

Figure 4: A grid of line graphs showing the results of a simulation comparing the Syrjala test to our proposed modified Syrjala test. In each line graph, the column name indicates the distributions (followed by the sample sizes) which are being tested. The blank underline (___) is a place holder for the row name. For example, the top left graph shows the number of significant tests (out of ten tests) on CSR realizations of 500 points with CSR realizations of 50 points.

to the data using five different numbers of rotations, namely 4, 6, 8, 10, and 36 rotations of the data within $360°$. Finally, for each different binning scenario (for the Syrjala test) or rotation number (for our modified Syrjala test) the respective test was applied ten times. The number of significant tests (p-value > 0.05) out of the ten iterations was recorded and a visualization of the simulation results is provided in Figure 4.

5.2 Simulation Results

Figure 4 displays a grid of line graphs which summarize the simulation study results. The first row of graphs compare cases when the null hypothesis is true (i.e., when both samples come from the same bivariate distributions). Realizations of CSR with 500 data points were compared to other realizations of CSR with 50, 100, 250, and 500 data points (as indicated in the respective columns from left to right). The underlined space (___) within each of the column names is a place holder for the row name. For example, the line graph in the top left of the grid is referred to as CSR 500 vs CSR 50.

Specifically, the CSR 500 vs CSR 50 graph shows the number of significant tests out of ten iterations for both the Syrjala and modified Syrjala tests across the different binning techniques (for the Syrjala test) or number of rotations (for our modified Syrjala test). Notice the Syrjala test failed to reject the null hypothesis for every iteration across the six binning techniques, whereas our modified Syrjala test rejected three tests (once when six rotations were employed within the test, once for eight rotations, and once for 36 rotations).

Looking at all of the comparisons of CSR vs CSR (all of the line graphs in the first row of Figure 4), we see that the Syrjala test rejected one out of 240 (ten iterations times six binning techniques times four sample size comparisons) tests. Our modified Syrjala test in contrast rejected nine out of 200 (ten iterations times five rotation levels times four sample size comparisons) tests. In other words, the Syrjala test rejected less than one percent of the tests, whereas our modified Syrjala test rejected 4.5% of the tests. Since we are testing at the 5% significance level, we should expect to see roughly 5% of tests reject the null hypothesis when it is actually true. Hence, our modified Syrjala test is more appropriately sensitive. This also confirms an observation made by Fuller et al. (2006): “In general, rejecting the null hypothesis of identical configurations is quite difficult with the Syrjala test”.

The remaining rows of graphs show comparisons between realizations of a CSR process with departures from CSR, i.e., when the null hypothesis is false. In the second row, realizations of CSR (with sample sizes of 500 points) were compared to realizations of a heterogeneous Poisson process called Center (with 50, 100, 250, and 500 sample points, respectively). Overall, the Syrjala test produced multiple non-significant tests depending on the binning technique and sample size. However, our modified Syrjala test detected significant differences uniformly across all of these cases regardless of the number of rotations.

Particularly, the CSR 500 vs Center 50 graph shows three non-significant tests for the Syrjala test when binning on a regular 5×5 grid, and six non-significant tests when binning with 25 random binning points. However, the Syrjala test produced significant results for 10×10 and 20×20 regular binning and for 100 and 400 random binning (each tested ten times). Additionally, when using 25 random binning points, the Syrjala test struggled to identify three, four, and five significant differences for each of the comparisons CSR 500 vs Center 100, CSR 500 vs Center 250, and CSR 500 vs Center 500, respectively. However, there are only significant Syrjala test results for all of the other binning techniques (5×5, 10×10, and 20×20 regular binning, as well as 100 and 400 random binning) for each of the CSR 500 vs Center 100, CSR 500 vs Center 250, and CSR 500 vs Center 500 comparisons. This suggests a dependence of the Syrjala test on the data aggregation step. It also suggests binning must be granular enough to reflect the deviations from CSR. In contrast, our modified Syrjala test detected significant differences uniformly across all of these CSR vs Center comparisons for any number of rotations.

In the third row, realizations of CSR were compared with departures from CSR called Repel. These comparisons provide an interesting case since both the Syrjala test and our modified Syrjala test struggled to indicate every significant difference across the different sample size comparisons. In the CSR 500 vs Repel 50 graph, the Syrjala test failed to detect any significant differences across all of the binning techniques. Whereas, our modified Syrjala test resulted in five, one, five, one, and four significant tests (each out of ten) for four, six, eight, ten, and 36 rotations, respectively.

In the CSR 500 vs Repel 100 graph, we begin to see positive association between the binning granularity and the Syrjala test. Notice when applying regular binning, the Syrjala test detected zero, one, and five significant tests for the 5×5, 10×10, and 20×20 regular grids, respectively. Similarly, when applying random binning, the Syrjala test detected zero, two, and five significant tests for the 25, 100, and 400 random binning points, respectively. In contrast, our modified Syrjala test produced nine, six, six, six, and eight significant tests for four, six, eight, ten, and 36 rotations, respectively. Hence, our modified Syrjala test is on average more powerful than the Syrjala test in this comparison.

In the remaining two comparisons (CSR 500 vs Repel 250 and CSR 500 vs Repel 500) our modified Syrjala test detected significant differences across all numbers of rotations. Whereas, the Syrjala test demonstrated a strong positive association with the binning granularity (as binning granularity increased, the number of significant tests also increased).

Overall, not only does row three in Figure 4 reinforce the observed dependence of the Syrjala test on the binning technique, but it also confirms that the Syrjala test places less emphasis on differences located near the center of the bounding region which was observed by McAdam et al. (2012). In contrast, our modified Syrjala test is in general more powerful in detecting these differences across all of the comparisons. Furthermore, while our modified Syrjala test averaged 3.2 significant tests in CSR 500 vs Repel 50, and seven significant tests in CSR 500 vs Repel 100, there is little indication in these cases of a dependence of our modified Syrjala test on the number of rotations.

In the remaining two rows where realizations of CSR are compared with the Corner and Right distributions, both tests were able to detect significant differences except for one case where the Syrjala test used a 5×5 regular binning grid. The case failed to detect the difference between a realization of CSR with 500 points and the Right distribution with 50 points. In general, however, these comparisons suggest that there are cases in which the Syrjala test and our modified Syrjala test still agree.

6. Application to the USU Posture Study

Along with the results of the simulation (in Section 5.2), we also conducted a comparison of p-values from the Syrjala test and our modified Syrjala test applied to data from the USU Posture Study (Symanzik et al., 2017, 2018). The study involved gathering eye-tracking data on subjects who were asked to determine the posture stability of a depicted person. Two groups of 20 subjects were labeled as the control and treatment groups depending on whether the subject had recent experience practicing yoga. One of the main questions of interest in the study is whether the two groups focus visually on different parts of the body when assessing another individual's stability. Figure 5 shows a comparison of two subjects' individual sample view points for a posture labeled as Posture 6.

Table 2 shows the p-values for the respective tests across the different data binning

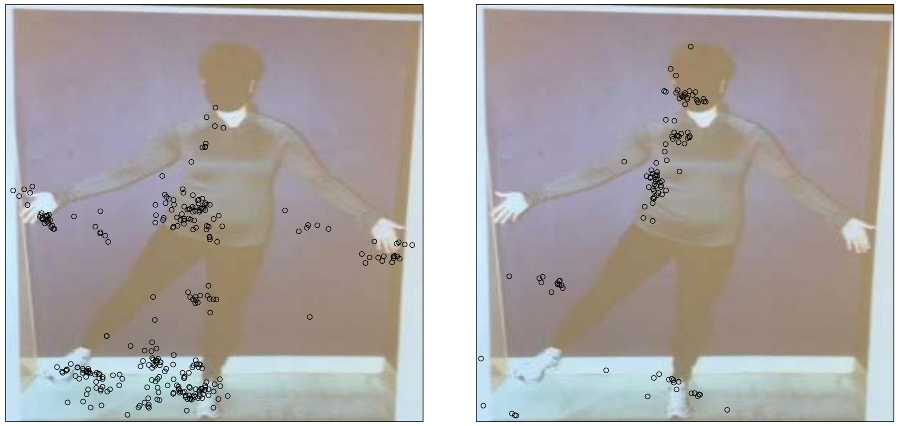

Figure 5: Comparison of Posture 6 of the USU Posture Study, showing viewing patterns for two subjects.

techniques and number of rotations (discussed in more detail in Section 5.1). For both the Syrjala test and our modified Syrjala test, 1000 permutations of the data were used to calculate the test statistics. Additionally, the Syrjala test and our modified Syrjala test were applied once for each binning technique or number of rotations, whereas the tests were iterated ten times in the simulation study for each binning technique. Notice the Syrjala test produces significant (bolded in Table 2) and non-significant p-values depending on the binning technique and number of bins. In contrast, our modified Syrjala test resulted in consistent significance across all numbers of data rotations within the test. This is consistent with the data since Figure 5 suggests differences in sample view points particularly around the head, upper torso, left knee, hands, and feet.

Table 2: Comparison of test results between the Syrjala test and our modified Syrjala test for Posture 6 viewing patterns from two subjects. The bolded values indicate significant results (p-values < 0.05).

	Syrjala Test						Modified Syrjala Test
	Regular Grid			Random Points			Rotations
Posture	5×5	10×10	20×20	25	100	400	4, 6, 8, 10, 36
6	0.245	0.060	**0.011**	0.186	**0.012**	**0.008**	**0.001**

7. Conclusions and Future Work

7.1 Conclusions

While cases of data aggregation have been used in the literature to achieve identical sampling locations necessary to apply the Syrjala test, through simulation (see Section 5) and application to real data (see Section 6), we demonstrated that the results of the Syrjala test depend on the binning of the data. We also confirmed an observation made by Fuller et al. (2006): "In general, rejecting the null hypothesis of identical configurations is quite difficult with the Syrjala test". Furthermore, we confirmed that the Syrjala test places less emphasis on differences centered in the bounding rectangle (McAdam et al., 2012).

In comparison, our modified Syrjala test produces results that are more appropriately sensitive and in general more powerful, especially in cases where the major differences exist in the center in the bounding rectangle (see Section 5.2). Our modified Syrjala test produces consistent results in our application to real data, and it is more generally applicable due to the removed restriction of identical sampling locations.

7.2 Future Work

While more computationally expensive, the simulation of our modified Syrjala test may be extended to higher numbers of rotations within 360° to ensure stability of results. Furthermore, additional versions of our modified Syrjala statistic could be examined including the following

$$\Psi_1^* = \sum_{r=1}^{R} \frac{1}{R} \Big\{ \sum_{i=1}^{n} \left[\Gamma_{1,r}^*(x_{1,i}, y_{1,i}) - \Gamma_{2,r}^*(x_{1,i}, y_{1,i})\right]^2 + \sum_{j=1}^{m} \left[\Gamma_{1,r}^*(x_{2,j}, y_{2,j}) - \Gamma_{2,r}^*(x_{2,j}, y_{2,j})\right]^2 \Big\}, \tag{8}$$

$$\Psi_2^* = \sum_{r=1}^{R} \frac{1}{R} \Big\{ \frac{nm}{(n+m)^2} \Big(\sum_{i=1}^{n} \left[\Gamma_{1,r}^*(x_{1,i}, y_{1,i}) - \Gamma_{2,r}^*(x_{1,i}, y_{1,i})\right]^2 + \sum_{j=1}^{m} \left[\Gamma_{1,r}^*(x_{2,j}, y_{2,j}) - \Gamma_{2,r}^*(x_{2,j}, y_{2,j})\right]^2 \Big) \Big\}, \text{ and} \tag{9}$$

$$\Psi_3^* = \sum_{r=1}^{R} \frac{1}{R} \Big\{ \frac{n}{(m+n)} \sum_{i=1}^{n} \left|\Gamma_{1,r}^*(x_{1,i}, y_{1,i}) - \Gamma_{2,r}^*(x_{1,i}, y_{1,i})\right| + \frac{m}{(m+n)} \sum_{j=1}^{m} \left|\Gamma_{1,r}^*(x_{2,j}, y_{2,j}) - \Gamma_{2,r}^*(x_{2,j}, y_{2,j})\right| \Big\}. \tag{10}$$

Equation (8) considers non-weighted sums of squared differences between the bivariate ecdf values from the two samples. Equation (9) is a construction similar to that proposed by Anderson (1962). Equation (10) calculates the weighted sums of absolute differences between the bivariate ecdf values.

8. Acknowledgments

We would like to thank our colleagues and fellow students at USU, particularly Dr. Breanna Studenka, Dr. Brennan Bean, Dr. Todd Moon, Joanna Coltrin, Kristina Casos, Melanie Athens, and Madison Hansen, for helpful discussions and for their efforts contributing to the USU Posture Study (Symanzik et al., 2017, 2018).

References

Anderson, T. W. (1962). On the Distribution of the Two-Sample Cramér-von Mises Criterion. *The Annals of Mathematical Statistics*, 33(3):1148–1159.

Benayas, J. M. R., de la Montaã, E., Pérez-Camacho, L., de la Cruz, M., Moreno-Mateos, D., L., P. J., Seoane, S. S., and Galván, I. (2010). Short-Term Dynamics and Spatial Pattern of Nocturnal Birds Inhabiting a Mediterranean Agricultural Mosaic. *Ardeola*, 57(2):303–320.

Berry, K. H., Coble, A. A., Yee, J. L., Mack, J. S., Perry, W. M., Anderson, K. M., and Brown, M. B. (2015). Distance to Human Populations Influences Epidemiology of Respiratory Disease in Desert Tortoises. *The Journal of Wildlife Management*, 79(1):122–136.

Berry, K. J., Johnston, J. E., and Mielke Jr, P. W. (2011). Permutation Methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 3(6):527–542.

Chetverikov, A., Kuvaldina, M., MacInnes, W. J., Jóhannesson, Ó. I., and Kristjánsson, Á. (2018). Implicit Processing During Change Blindness Revealed with Mouse-Contingent and Gaze-Contingent Displays. *Attention, Perception, & Psychophysics*, 80(4):844–859.

Chiu, S. N. and Liu, K. I. (2009). Generalized Cramér-von Mises Goodness-of-Fit Tests for Multivariate Distributions. *Computational Statistics & Data Analysis*, 53(11):3817–3834.

Conover, W. J. (1998). *Practical Nonparametric Statistics (3rd Edition)*. Wiley, New York, NY.

Cramér, H. (1928). On the Composition of Elementary Errors: First Paper: Mathematical Deductions. *Scandinavian Actuarial Journal*, 1928(1):13–74.

Fuller, T., Munguía, M., Mayfield, M., Sánchez-Cordero, V., and Sarkar, S. (2006). Incorporating Connectivity into Conservation Planning: A Multi-Criteria Case Study from Central Mexico. *Biological Conservation*, 133(2):131–142.

Kolmogorov, A. (1933). Sulla Determinazione Empirica di una Legge di Distributione. *Giornale dell'Istituto Italiano degli Attuari*, 1933(4):83–91. (In Italian).

McAdam, B. J., Grabowski, T. B., and Marteinsdóttir, G. (2012). Testing for Differences in Spatial Distributions from Individual Based Data. *Fisheries Research*, 127(1):148–153.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`.

Symanzik, J., Li, C., Zhang, B., Studenka, B. E., and McKinney, E. (2017). Eye-Tracking in Practice: A First Analysis of a Study on Human Postures. In *2017 JSM Proceedings.*, Alexandria, VA. American Statistical Association. (CD).

Symanzik, J., McKinney, E., Studenka, B. E., Bean, B., Athens, M., and Hansen, M. (2018). Eye-Tracking in Practice: Results from a Study on Human Postures. In *2018 JSM Proceedings.*, Alexandria, VA. American Statistical Association. (CD).

Syrjala, S. E. (1996). A Statistical Test for a Difference Between the Spatial Distributions of Two Populations. *Ecology*, 77(1):75–80.

von Mises, R. (1928). *Wahrscheinlichkeit Statistik und Wahrheit*. Springer-Verlag, Wien.