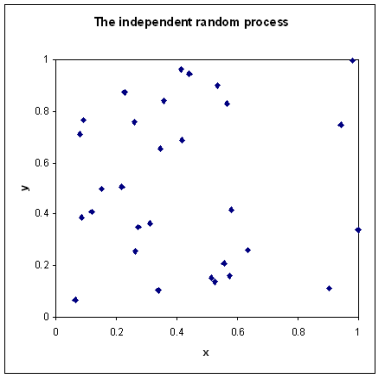
Geography 586

Project 3 - Understanding Random Spatial Processes

Bruce Kinner

1. *Examples of patterns generated by the independent random process*



The human mind naturally seeks structure from the chaos of existence. A plot of randomly generated data will trigger unconscious axiomatic thought in an attempt to stratify and order the visual representation. Many of the randomly generated data graphs do not appear random at all. Simple stories can be easily conjured to bestow meaning. For example, after contemplating the graph on the left for a few minutes, I invented a whole story to explain the graph. Clearly, the graph shows the grades (x) everyone (y) received on quizzes in a graduate geospatial statistics class. Of course, most students failed utterly while a few nerds aced the questions. After generating a few dozen more random graphs, I found another one that fit the story - the re-test! This time everyone actually studied and did much better with over half the class passing the exam. The point is if a story can be invented for data known to be generated reandomly, caution should be taken with data of unknown processes!

The graph on the right is one where no story came to mind. The points are spread throughout the map and while there are a few small clusters no trend is identifiable. This shows how fine a line it can be. If the point at approximately (0.5,0.4) is removed a large hole in created in the middle of the graph which would make the graph appear much less random.

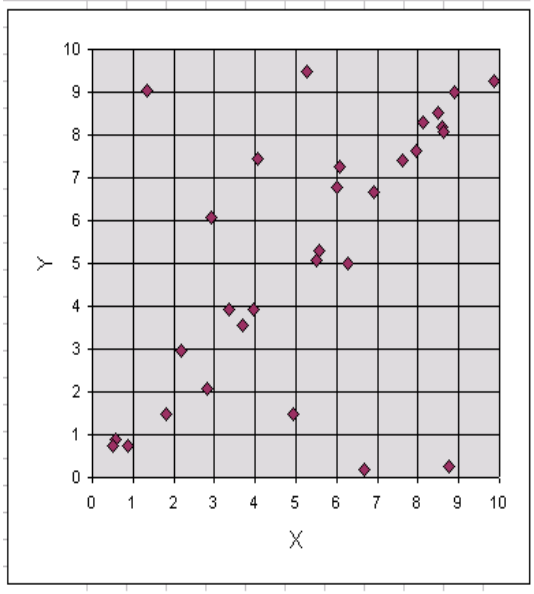
The counter-intuitive conclusion is clustering and empty regions are a natural part of randomly generated data. A certain percentage of random graphs like these will have patterns that strongly suggest trends simply because, just like rolling multiple dice at once, there is a chance of receiving all the same results with each random event in the set.

2. *Examples of patterns generated by the 'Inhomogeneous Poisson' spreadsheet.*

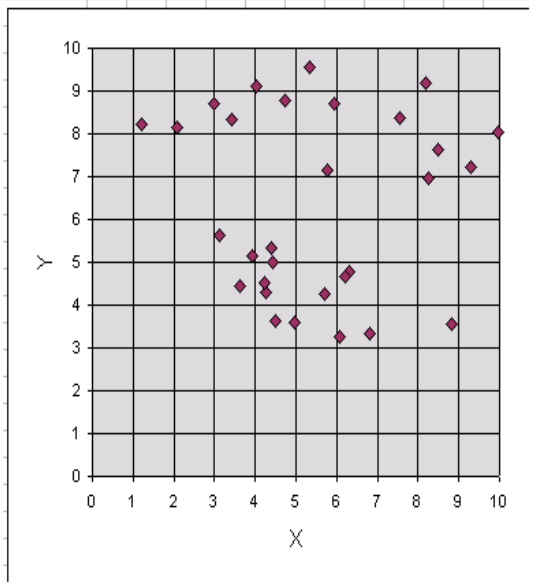
The graphs at the top of the next page show how a random process can be controlled by weighting. By creating a heavy weight to the initial values, the random plot can be controlled to create data with an expected outcome. In the graph on the left, the slope to the line is to be expected from the weighted values - surely x and y are correlated!

The graph on the right is more subtle. A strong but not overwhelming weight was given to the center section but the weights for the top and right of the graph are still significant. In this particular incarnation, the top has randomly received more points than the right even though the chances for those areas are the same. The clustering and empty regions shown in the right graph below are similar to the ones in the first two IRP graphs. This shows data generated by a Inhomogeneous Poisson method can appear similar to ones generated by the independent random process (IRP).

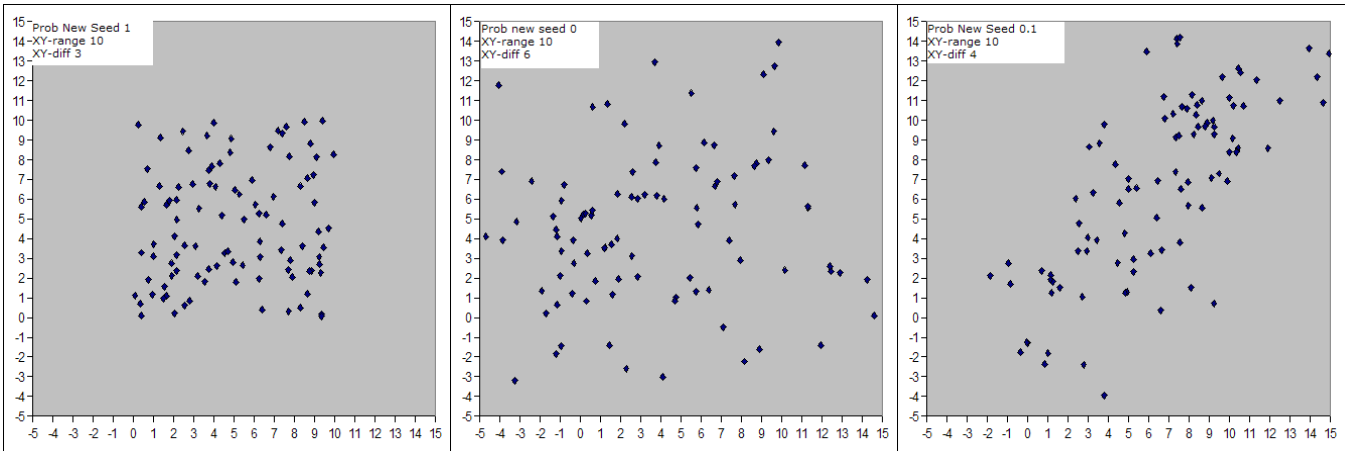
	Column									
Row	0	1	2	3	4	5	6	7	8	9
9	10	11	12	13	14	15	16	17	18	200
8	9	10	11	12	13	14	15	16	200	18
7	8	9	10	11	12	13	14	200	16	17
6	7	8	9	10	11	12	200	14	15	16
5	6	7	8	9	10	200	12	13	14	15
4	5	6	7	8	200	10	11	12	13	14
3	4	5	6	200	8	9	10	11	12	13
2	3	4	200	6	7	8	9	10	11	12
1	2	200	4	5	6	7	8	9	10	11
0	200	2	3	4	5	6	7	8	9	10



	Column									
Row	0	1	2	3	4	5	6	7	8	9
9	10	11	12	13	14	15	16	17	18	1
8	9	10	11	12	13	14	15	16	2	18
7	8	9	10	11	12	13	14	3	16	17
6	7	8	9	40	40	40	40	14	15	16
5	6	7	8	40	80	80	40	13	14	15
4	5	6	7	40	80	80	40	12	13	14
3	4	5	6	40	40	40	40	11	12	13
2	3	4	3	6	7	8	9	10	11	12
1	2	2	4	5	6	7	8	9	10	11
0	1	2	3	4	5	6	7	8	9	10



3. *Examples of patterns generated by the 'Interaction effects' spreadsheet.*



The patterns generated by the 'Interaction effects' are the most controllable but still exhibit random behavior. It is possible to generate data exactly like the IRP. By setting the probability of the 'new seed' to 1, each point is randomly placed within the XY-range (0-10 here instead of 0-1 in the IRP). At the other extreme, a 'new seed' of 0 means all the subsequent points after the first will be based on existing point. Even so, most of the graphs generated in this manner appear as random as the ones generated with the IRP (as shown in the middle graph above) - a few natural clusters and some empty regions.

The scale of the axis becomes important. The ratio of XY-diff to XY-range controls the area the points will cover. A small ratio means the points will be tightly packed on the number line and cover a small area while a large ratio means the point will be widely spread. However, the relative appearance and randomness of the graph is the same. Finally, with a 'new seed' value which allows a few random points to be place it is possible but unlikely to generate a graph (shown at right) similar to the controlled slope graph in part 2.

The total new seeds generated is not really related to the number of visual clusters in the final output. Often there is enough overlap even after two or three new seeds for the data create distinguishable clusters. Again the ratio of XY-diff to XY-range is important. Very low ratios (0.2 or less) can (but not always) produce very clear clusters especially if the probability of a new seed is low.