Off the Beaten Path Tutorial: Stochastic Processes and Simulations – Volume 1

Vincent Granville, Ph.D.
January 2022
www.datashaping.com

Note: External links (in blue) and internal references (in red) are clickable throughout this document. Keywords highlighted in orange are indexed; those in red are both indexed and in the glossary section.

Contents

List of Figures 3

List of Symbols and Abbreviations 3

1 About this Textbook 3
- 1.1 Brief Presentation . . . 3
- 1.2 Target Audience . . . 4
- 1.3 List of Exercises . . . 4
- 1.4 About the Author . . . 5

2 Poisson-binomial or Perturbed Lattice Process 5
- 2.1 Definitions . . . 6
- 2.2 Counting Measure and Interarrival Times . . . 7
- 2.3 Limiting Distributions, Speed of Convergence . . . 8
- 2.4 Properties of Stochastic Point Processes . . . 9
  - 2.4.1 Stationarity . . . 9
  - 2.4.2 Ergodicity . . . 10
  - 2.4.3 Independent Increments . . . 10
  - 2.4.4 Homogeneity . . . 10
- 2.5 Transforming and Combining Multiple Point Processes . . . 10
  - 2.5.1 Marked Point Process . . . 10
  - 2.5.2 Rotation, Stretching, Translation and Standardization . . . 11
  - 2.5.3 Superimposition and Mixing . . . 11
  - 2.5.4 Hexagonal Lattice, Nearest Neighbors . . . 12

3 Applications 13
- 3.1 Modeling Cluster Systems in Two Dimensions . . . 14
  - 3.1.1 Generalized Logistic Distribution . . . 15
  - 3.1.2 Illustrations . . . 16
- 3.2 Infinite Random Permutations with Local Perturbations . . . 17
- 3.3 Probabilistic Number Theory and Experimental Maths . . . 18
  - 3.3.1 Poisson Limit of the Logistic-binomial Distribution, with Applications . . . 18
  - 3.3.2 Perturbed Version of the Riemann Hypothesis . . . 21

4 Statistical Inference, Machine Learning, and Simulations 22
- 4.1 Model-free Tests and Confidence Intervals . . . 22
  - 4.1.1 Empirical Distributions . . . 23
  - 4.1.2 A New Test of Independence . . . 23
- 4.2 Estimation of Core Parameters . . . 23
  - 4.2.1 Intensity and Scaling Factor . . . 23
  - 4.2.2 Model Selection to Identify F . . . 24
  - 4.2.3 Theoretical Values Obtained by Simulations . . . 25
- 4.3 Hard-to-Detect Patterns and Model Identifiability . . . 26
- 4.4 Spatial Statistics, Nearest Neighbors, Clustering . . . 28
  - 4.4.1 Stochastic Residues . . . 28
  - 4.4.2 Inference for Two-dimensional Processes . . . 29

4.4.3 Clustering Using GPU-based Image Filtering . . . 32
4.4.4 Black-box Elbow Rule to Detect Outliers and Number of Clusters . . . 33
4.5 Boundary Effect . . . 36
4.5.1 Quantifying some Biases . . . 37
4.5.2 Extreme Values . . . 38
4.6 Poor Random Numbers and Other Glitches . . . 40
4.6.1 A New Type of Pseudo-random Number Generator . . . 41

5 Theorems **41**
5.1 Notations . . . 41
5.2 Link between Interarrival Times and Point Count . . . 42
5.3 Point Count Arithmetic . . . 42
5.4 Link between Intensity and Scaling Factor . . . 42
5.5 Expectation and Limit Distribution of Interarrival Times . . . 43
5.6 Convergence to the Poisson Process . . . 44
5.7 The Inverse or Hidden Model . . . 45
5.8 Special Cases with Exact Formula . . . 46
5.9 Fundamental Theorem of Statistics . . . 47

6 Exercises, with Solutions **47**
6.1 Probability Distributions, Limits and Convergence . . . 47
6.2 Features of Poisson-binomial Processes . . . 51
6.3 Lattice Networks, Covering Problems, and Nearest Neighbors . . . 52
6.4 Miscellaneous . . . 56

7 Source Code, Data and Excel Spreadsheets **57**
7.1 Interactive Spreadsheets . . . 59
7.2 Source Code: Point Count, Interarrival Times . . . 59
7.2.1 Compute $\mathrm{E}[N(B)], \mathrm{Var}[N(B)]$ and $P[N(B) = 0]$ . . . 60
7.2.2 Compute $\mathrm{E}[T], \mathrm{Var}[T]$ and $\mathrm{E}[T^r]$ . . . 61
7.2.3 Produce random deviates for various F's . . . 62
7.2.4 Compute $F(x)$ for Various F's . . . 62
7.3 Source Code: Radial Cluster Simulation . . . 62
7.4 Source Code: Nearest Neighbor Distances . . . 63
7.5 Source Code: Detection of Connected Components . . . 67
7.6 Source Code: Visualizations, Density Maps . . . 68
7.6.1 Visualizing the Nearest Neighbor Graph . . . 68
7.6.2 Clustering and Density Estimation via Image Filtering . . . 69

Glossary **72**

References **74**

Index **77**