Applications of Stochastics — Simulation projects

Gábor Pete
http://www.math.bme.hu/~gabor

1. **Aldous' theorem.** Show that the vector of the two largest cluster sizes $C_1(n), C_2(n)$ in the critical Erdős-Rényi graph $G(n, 1/n)$, scaled by $n^{2/3}$, converges in distribution to the vector of the two longest excursions of a Brownian motion with parabolic drift, $B_t - t^2/2$, away from its running minimum (see PGG Theorem 12.23). The process $(B_t)_{t\geq 0}$ should be simulated as the limit of $X_{nt}/\sqrt{n}$ as $n \to \infty$, where X_i is simple symmetric random walk on $\mathbb{Z}$.

2. **Random walk in random environment.** Let $\{p_i : i \in \mathbb{Z}\}$ be an iid sequence with $p_i \in (0,1)$. Fix this random environment, then consider the random walk

$$\mathbf{P}[\, X_{n+1} = i+1 \mid X_n = i\,] = 1 - \mathbf{P}[\, X_{n+1} = i-1 \mid X_n = i\,] = p_i\,.$$

   (a) Let p_i be 1/3 or 2/3, with probability 1/2 each. Find a deterministic sequence a_n such that X_n/a_n seems to be converging in distribution to a non-degenerate variable.

   (b) Find a distribution with $\mathbf{E}p_i = 1/2$ such that X_n is transient.

   (c) In both cases, draw pictures of the space-time trajectories.

3. **A random walk in Manhattan.** In Manhattan, all streets are one-way. So, for each infinite line of $\mathbb{Z}^2$, flip a fair coin, orienting it one way or the other. Given this random environment, let X_n be the random walk that, at each corner, chooses one of the two possibilities (continuing straight or turning, respecting the one-way direction) with probability 1/2. Is this walk recurrent? How far is typically X_n from the origin, for large n? Draw pictures of the trajectory.

4. **Random walk in a changing random environment.** Consider critical dynamical bond percolation on the $n \times n$ torus $(\mathbb{Z}/n\mathbb{Z})^2$: at the beginning, each edge is open with probability 1/2, independently, then at each time step, one edge is chosen at random and its status is flipped. Now consider a particle performing a random walk in this changing maze with "infinite speed": that is, it is always uniformly distributed in its current cluster. When a cluster gets divided into two, the particle choses one of the resulting clusters, with probabilities proportional to the sizes. Starting from the cluster of the origin, how many steps are needed for the particle to be approximately uniformly distributed on the torus?

5. **PageRank for Barabási-Albert.** How does the score of a vertex correlate with time? E.g., arrival time of first ranked vertex goes to infinity with growth of graph?

6. **Random genetic drift drives a population towards genetic uniformity.** Consider the Wright-Fisher model, as follows. A certain gene can have two alleles, A and B. At the beginning, the two alleles are represented equally in the gene pool given by N diploid individuals: there are altogether N copies of a A and N copies of B. In the next generation, we again have N individuals, with each of their altogether $2N$ genes drawn independently at random from all the genes in the old generation. And so on, repeated forever.

   (a) How many generations does it typically take to eradicate one of the alleles from the gene pool?

(b) Now assume that, in each generation, each individual may go dormant, independently with probability λ/N, some $\lambda \in (0,\infty)$ fixed, and stays dormant for an independent time ξ with distribution $\mathbf{P}[\,\xi \geq t\,] = t^{-\beta}$, $t = 1,2,3,\ldots$, some $\beta > 0$. When D individuals are dormant, then the reproduction is like before, just with the $N-D$ non-dormant individuals participating. When an individual wakes up, it will take part in the reproduction, and thus may re-introduce a seemingly extinct allele. For what values of λ and β is the time scale to get complete uniformity significantly larger than before?

7. **Positive overshoots with negative drift.** Consider a random walk $S_n = X_1 + \cdots + X_n$ on $\mathbb{R}$, with iid increments satisfying $\mathbf{E}X_i < 0$, but $\mathbf{P}[\,X_i > 0\,] > 0$, moreover, with $\mathbf{E}(X_i \vee 0)^2 = \infty$. (In particular, the size-biased version of $X_i \vee 0$ exists, but has infinite expectation.) Let $T := \inf\{n > 0 : S_n > 0\}$, where the infimum is defined to be infinite if the set is empty.

   (a) Does it seem to be always true that $\mathbf{E}[\,S_T \mid T < \infty\,] < \infty$?

   (b) Does it seem to be always true that $\mathbf{E}[\,S_T \mid T < \infty\,] = \infty$?

8. **Bootstrap percolation.** In the $n \times n$ box in $\mathbb{Z}^2$, start with an i.i.d. Bernoulli(p) set of occupied vertices. Then, at each round, a vertex becomes occupied if at least 2 of its 4 neighbours are occupied; this is repeated until there are no changes in a round.

   (a) Estimate the critical value $p_c(n)$ for the initial occupation density p for which the probability that every vertex becomes eventually occupied is 1/2.

   (b) Around the critical density, take an instance when complete occupation happens, and make a picture of the occupation process: let the colour of a site (or, for better visibility, of a unit square) depend on the round in which it got occupied.

9. **Liquid crystal.** In $\mathbb{R}^2/(n\mathbb{Z})^2$, the 2-dimensional torus of side length n, let $X_1, X_2, \ldots$ be iid uniform random points. From each X_i iteratively, draw a unit vector at a uniform random angle, unless it intersects some previously drawn vector. Do this until we have n^2 vectors drawn.

   (a) How many tries are needed typically?

   (b) In a typical subsquare of side-length m, there are of order m^2 vectors. One can say that they are pointing roughly in the same direction (there is long range order in this subsquare) if their vector sum has length of order m^2. What is the largest $m = m(n)$ for which a typical subsquare has long range order?

   (c) Make pictures.

10. **Gaussian copula.** Consider the following data from the last 100 days for the prices of a pair of stocks:

    {199.183, 198.731}, {199.974, 199.734}, {198.084, 198.307}, {199.132, 200.579}, {198.995, 200.155}, {199.744, 199.44}, {199.546, 198.39}, {199.755, 200.776}, {198.47, 199.096}, {198.662, 199.675}, {189.774, 189.307}, {186.628, 186.131}, {196.473, 198.316}, {199.803, 199.613}, {197.333, 198.765}, {198.407, 199.52}, {199.989, 200.138}, {196.261, 196.598}, {199.866, 201.095}, {196.152, 195.168}, {200.021, 199.419}, {199.622, 198.419}, {200.605, 200.932}, {196.332, 194.418}, {193.769, 196.125}, {196.958, 197.247}, {198.648, 198.961}, {199.039, 199.532}, {198.371, 198.722}, {197.122, 200.102}, {196.644, 198.725}, {199.822, 199.674}, {199.112, 199.773}, {197.595, 196.657}, {199.663, 197.82}, {199.039, 199.135}, {196.899, 198.705}, {199.176, 200.07}, {198.626, 200.604}, {199.48, 200.255}, {195.652, 197.964}, {199.708, 199.213}, {198.009, 198.869}, {199.743, 199.869}, {196.87, 200.09}, {193.913, 192.382}, {196.284, 198.334}, {199.07, 200.245}, {198.899, 200.216}, {200.407, 198.075}, {199.626, 200.985}, {199.278, 197.229}, {199.512, 200.966}, {190.633, 192.106}, {198.982, 198.297}, {200.74, 200.235}, {199.366, 198.7}, {200.311, 200.237}, {199.723, 199.197}, {195.653, 197.154}, {190.626, 189.285}, {199.477, 199.724}, {199.296, 199.29}, {142.269, 144.896}, {198.028, 197.95}, {198.072, 197.308}, {198.153, 199.564}, {190.066, 188.593}, {200.105, 200.592}, {198.656, 200.201}, {199.411, 198.112}, {199.17, 197.28}, {200.371, 200.146}, {198.712, 198.329}, {198.956, 200.89}, {200.183, 196.989}, {187.394, 187.588}, {198.15, 199.422}, {173.914, 174.935}, {197.05, 198.765}, {199.175, 199.964}, {198.341, 198.239}, {197.813, 196.851}, {200.743, 199.522}, {184.203, 185.063}, {199.543, 196.427}, {198.676, 198.976}, {198.362, 198.797}, {199.965, 198.247}, {199.082, 198.926}, {201.179, 199.302}, {198.334, 198.182}, {188.417, 186.537}, {198.011, 199.522}, {201.118, 200.}, {198.235, 194.815}, {200.166, 198.914}, {198.035, 198.593}, {199.276, 199.479}, {200.556, 197.852}.

(a) Calculate the sample mean vector μ and covariance matrix Σ for this data.

(b) Assuming that the distribution is bivariate normal, with the parameters (μ, Σ) just obtained, make a random sample how the next 100 days may look like.

(c) Estimate the marginal distributions of the data, then using the Gaussian copula with parameters (μ, Σ), make a random sample for the next 100 days.

(d) Vice versa, calculate the sample copula of the data, then assume that the marginals are normal, with marginal parameters obtained above, make a random sample for the next 100 days.

(e) Now use the marginals and the copula obtained from the data, and make a random sample for the next 100 days.

(f) Plot all the data: (1) the original; (2) bivariate normal; (3) estimated marginals, Gaussian copula; (4) estimated copula, Gaussian marginals; (5) estimated copula, estimated marginals. How similar are these to each other?

(g) We go bankrupt in the future if both prices go below 0. For which model does this seem to be the most likely? (Of course, since the minimum number in the entire data is 142.269, it's not really possible to estimate this probability. I'm just asking for simple-minded intuition, which is what many traders would also rely on.)