# Counting self-avoiding paths in the plane with dual variables

This statistical mechanics model grew out of discussions with Persi Diaconis on a method devised by Donald Knuth[1] for counting self-avoiding paths in the plane. Knuth's method should *not* be used for this problem, and self-avoiding paths mostly serve to illustrate how seemingly reasonable methods can fail spectacularly. Nevertheless, the article is entertaining and instructive, even 40 years after publication.

Knuth was interested in reliably estimating the number of self-avoiding paths within an $m\times n$ rectangle that step along the edges of the integer lattice and have endpoints at diagonally opposite corners. His (doomed) method, *sequential importance sampling*, is closer to a strategy a statistician would propose than something inspired by statistical mechanics.

An example of a statistical-mechanics inspired method is to define a Hamiltonian proportional to the *excess path length* relative to monotonic paths, of which there are $\binom{m+n}{m}$. At zero temperature we only have monotonic paths (whose number we know), while at infinite temperature we have all self-avoiding paths (whose number we don't know). The statistical mechanics trick, called the "integration method", relates the two numbers by integrating the derivative of the entropy (the heat capacity) between the two limiting temperatures. While this method would work, it has a *scaling* problem that the far superior (and less obvious) method described below does not.

The main idea is to use variables that are "dual" to the sites visited by the path. These live on the elementary squares of the lattice (their centers). An edge that joins adjacent sites of the path intersects a dual edge that joins adjacent squares. This kind of duality does not generalize to self-avoiding paths in higher dimensions.

The expressive power of the dual variables is explained in Figure 1 (see caption). The rectangle of the walk is surrounded by a frame of squares on all sides, so there is a binary dual variable for every square within an $(m+2)\times(n+2)$ rectangle. The variables in the frame are static and fix the boundary conditions of the path. Only the variables inside the original $m\times n$ rectangle take different values and these span all $2^{mn}$ possibilities.

Figure 1 illustrates the case where the frame variables are set for paths that run between diagonal corners. To get paths *without* endpoints one would instead set all the frame variables to the same value. We get self-avoiding paths when the number of edge-connected sets of like-valued dual variables is the smallest possible: 2.

[1] D. Knuth, *Mathematics and Computer Science: Coping with Finiteness*, Science **194**, 1235-1242 (1976).

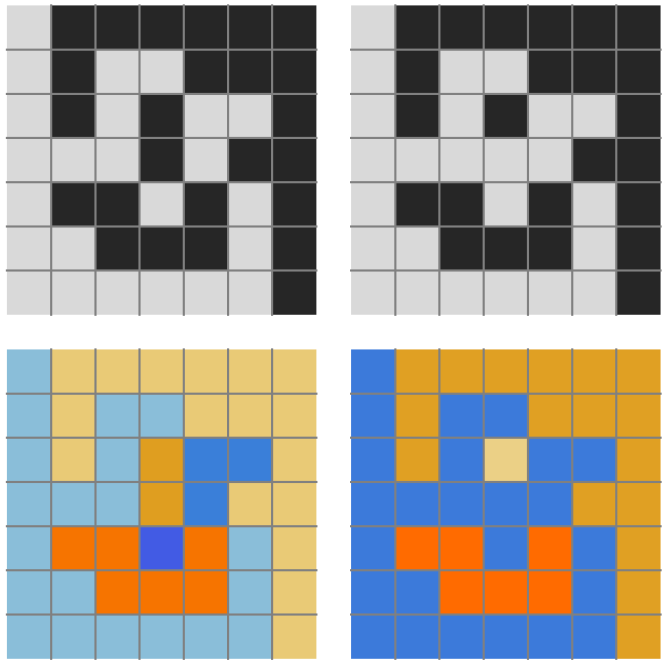

Figure 1: *Top*: Binary dual variables for the case $m = n = 5$; variables in the surrounding frame are set for generating paths having endpoints at the upper left and lower right corners. *Bottom*: Edge-connected sets of squares that have the same binary value in the top panel are colored with distinct colors. *Left/Right*: The left and right panels are related by flipping the value of the central dual variable. The excess number of connected components goes from $x(c) = 4$ (left) to $x(c') = 2$ (right).

Let $x(c)$ be the excess number of connected components in dual-variable configuration $c$. Define a partition function that penalizes this quantity:

$$Z(\beta) = \sum_c e^{-\beta x(c)}. \tag{1}$$

Since $Z(0) = 2^{mn}$, and $Z(\infty)$ is the number of self-avoiding paths, we can relate the logarithms of these, the entropy, by the integral of an expectation value:

$$s = \log Z(\infty) = mn \log 2 - \int_0^\infty e(\beta) d\beta, \tag{2}$$

where

$$e(\beta) = -\frac{d}{d\beta} \log Z = E(x(c)). \tag{3}$$

From this we see the main advantage of the dual variable method: the entropy of the reference state, now at $\beta = 0$, has the same scaling behavior with area as the self-avoiding paths.

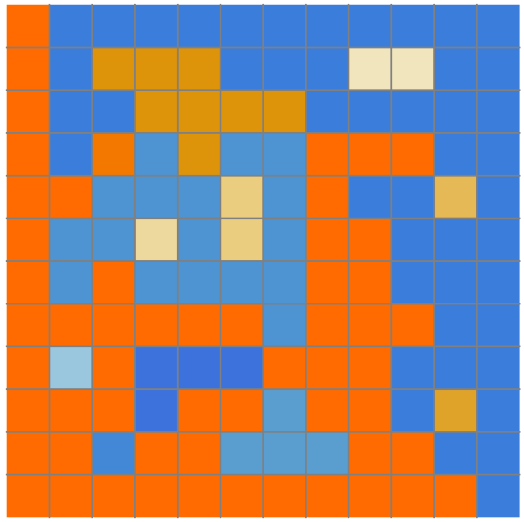

Figure 2: Typical configuration of dual variables for paths in the $10 \times 10$ square when $\beta = 0$.

Figures 2-4 show typical configurations for the case $m = n = 10$ as $\beta$ is increased from 0 to 1 and then to 6. The average number of excess connected components decays exponentially with $\beta$ and is quite small already for $\beta > 6$ as shown in Figure 5. The area under this curve, when subtracted from $10 \times 10 \log 2$, gives the entropy for paths in the $10 \times 10$ square.

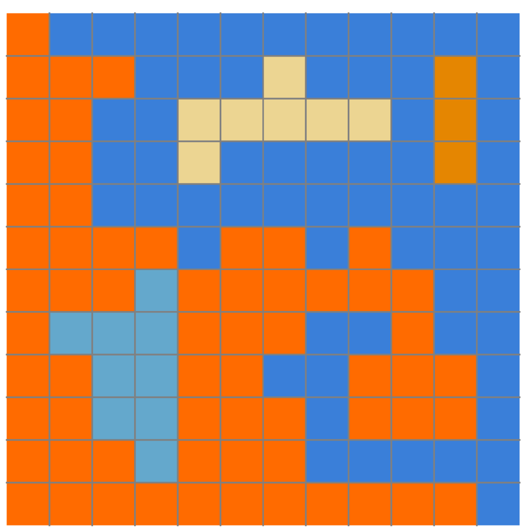

Figure 3: Same as Figure 2 but for $\beta = 1$. In this configuration $x(c) = 3$.

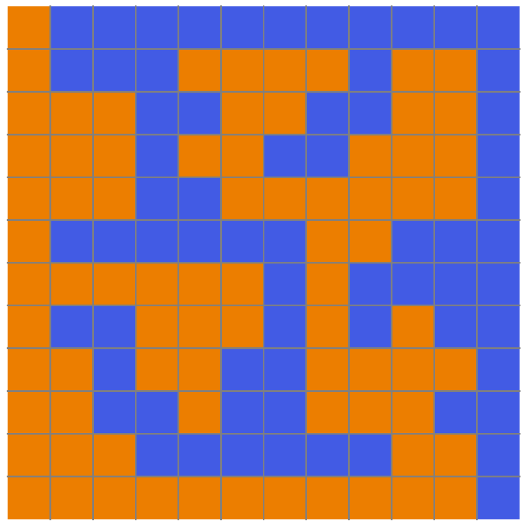

Figure 4: Same as Figure 2 but for $\beta = 6$. This configuration has $x(c) = 0$ and corresponds to a self-avoiding path.

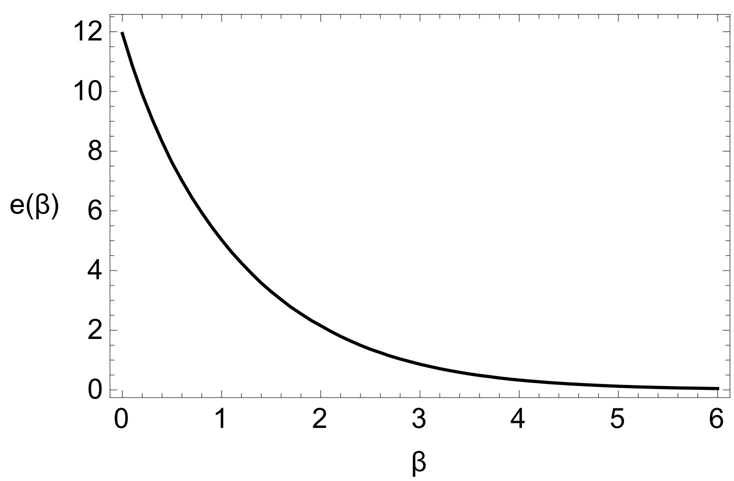


Figure 5: Average of the number of excess connected components, Eq. (3), for paths in the $10 \times 10$ square.