Distributed Density Estimation

Amartya Shankha Biswas *

1 Domino Tiling

We will consider the problem of tiling a square grid with dominos. This problem has a long histor and various importtant applications in statistical physics. Specifically, we will focus on the local generation of domino tilings from the uniform distribution.

1.1 $2 \times n$ Domino Tiling

The simplest version of the problem is one where we are given a $2 \times n$ grid (Figure **??**. The queries will be as an index, and the generator should report the orientation of the domino at the i^{th} position in the grid.

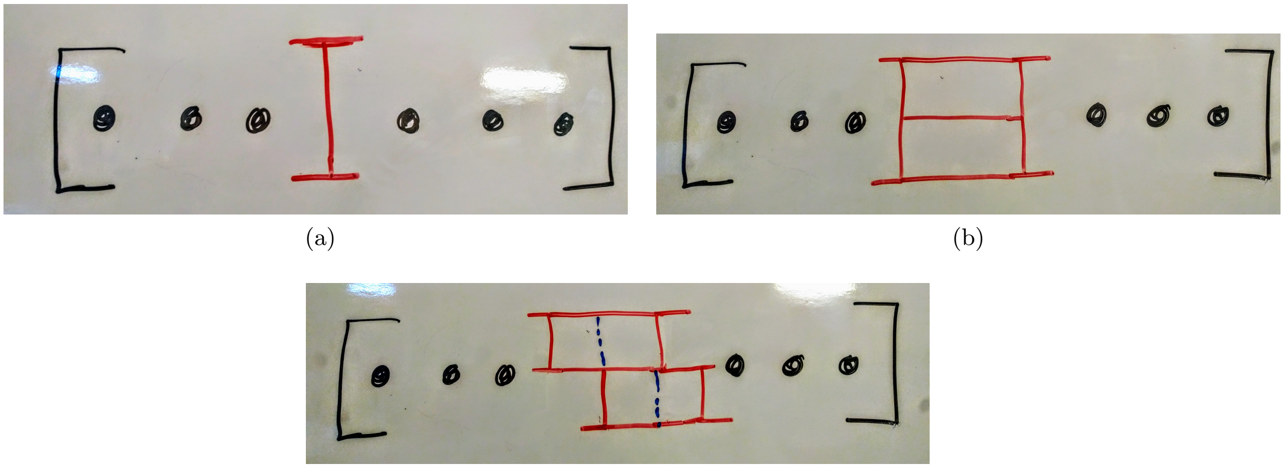
It is a well known result that the number of tilings of a $2 \times n$ grid is exactly F_n. cite

To aid with generalization, we will instead allow the generator to respond with the splitting boundary of the current tiling instead. For example, in Figure 1a, the boundary is a vertical line at the specified position. In Figure 1b, the boundary is horizontal, indicating that there are two horizontal dominos at that location. Note that Figure 1c is impossible for a $2 \times n$ grid. It should be clear that this query model is equivalent.

Now, consider a query to the location i, such that all positions between $i - a$ and $i + b$ have not been queriesd so far. So, there is a blank $2 \times (a + b)$ size sub-grid that we have to sample from. Let us consider the number of possible tilings resulting from each possible splitting boundary.

1. Vertical Boundary – This indicates that we divide the region into two sub-grids with sizes $2 \times a$ and $2 \times b$. So, the total number of possible tilings is exactly $F_a \cdot F_b$.
2. Horizontal Boundary – This indicates that we divide the region into two sub-grids with sizes $2 \times (a - 1)$ and $2 \times (b - 1)$. So, the total number of possible tilings is exactly $F_{a-1} \cdot F_{b-1}$.

*MIT, Cambridge MA 02139. E-mail: `asbiswas@mit.edu`.



(a) (b)

(c)

Figure 1: Caption for this figure with two images

So the probabilities are computed as $\frac{F_a \cdot F_b}{F_a \cdot F_b + F_{a-1} \cdot F_{b-1}}$ and $\frac{F_{a-1} \cdot F_{b-1}}{F_a \cdot F_b + F_{a-1} \cdot F_{b-1}}$. Now, we face the issue of approximating these fractions. If either of the values a or b are less than $\Theta(\sqrt{n})$, then we can compute the exact value of the corresponding F_a or F_b. Otherwise, we use Lemma **??** to approximate $F_a = \phi \cdot F_{a-1}$ and $F_b = \phi \cdot F_{b-1}$. So, the probability of the vertical boundary becomes

$$\frac{F_a \cdot F_b}{F_a \cdot F_b + F_{a-1} \cdot F_{b-1}} = \frac{\phi^2}{\phi^2 + 1}$$

Similarly, the probability of a horizontal split with a top and bottom domino becomes $1/(\phi^2 + 1)$. Note that this also determines the two adjacent boundaries.

The only information we needed to make this query was the extent of the un-queried interval $[i - a, i + b]$. We can use any standard data-structure that allows insertion in positions $\{1, 2 \cdots, n+1\}$, and provides successor and predecessor queries.

Here's we can be fancy and use Van-Emde-Boas trees to get a $\mathcal{O}(\log \log n)$ query time. However, in some cases, the exact value of a Fibonacci number still needs to be computed, and this takes $\mathcal{O}(\log n)$ time. The faster queries only work when the new query is "far enough" ($\mathcal{O}(\log n)$ distance) away from all previous queries.

References

A Dyck Path Generator

A.1 Assumptions

fix

- $d < c \cdot \sqrt{S} \log n$
- $k < c \cdot \sqrt{S} \log n \implies U - D < c \cdot \sqrt{S} \log n$
- $k' < c \cdot \sqrt{S} \log n$
- $S > \log^2 n \implies \sqrt{S} \log n < S$

Lemma 1. *For $x < 1$ and $k \geq 1$,*

$$1 - kx < (1-x)^k < 1 - kx + \frac{k(k-1)}{2} x^2.$$

Lemma 2. $D_{left} \leq c_1 \frac{k \cdot \log n}{\sqrt{S}} \cdot \binom{S}{D-d}$ *for some constant c_1.*

Proof. This involves some simple manipulations.

$$\begin{aligned}
D_{left} &= \binom{S}{D-d} - \binom{S}{D-d-k} && (1)\\
&= \binom{S}{D-d} \cdot \left[1 - \frac{(D-d)(D-d-1)\cdots(D-d-k+1)}{(S-D-d+k)(S-D-d+k-1)\cdots(S-D-d+1)}\right] && (2)\\
&\leq \binom{S}{D-d} \cdot \left[1 - \left(\frac{D-d-k+1}{S-D+d+k}\right)^k\right] && (3)\\
&\leq \binom{S}{D-d} \cdot \left[1 - \left(\frac{U+d+k-(U-D+d+k-1)}{U+d+k}\right)^k\right] && (4)\\
&\leq \binom{S}{D-d} \cdot \left[1 - \left(\frac{U+d+k-\mathcal{O}(\log n\sqrt{S})}{U+d+k}\right)^k\right] && (5)\\
&\leq \Theta\left(\frac{k \log n}{\sqrt{S}}\right) \cdot \binom{S}{D-d} && (6)
\end{aligned}$$

□

Lemma 3. $D_{right} < c_2 \frac{k' \cdot logn}{\sqrt{S}} \cdot \binom{S}{U-d}$ *for some constant c_2.*

Proof.

$$D_{right} = \binom{S}{U-d} - \binom{S}{U-d-k'} \tag{7}$$
$$= \binom{S}{U-d} \cdot \left[1 - \frac{(U-d)(U-d-1)\cdots(U-d-k'+1)}{(S-U-d+k')(S-U-d+k'-1)\cdots(S-U-d+1)}\right] \tag{8}$$
$$\leq \binom{S}{U-d} \cdot \left[1 - \left(\frac{U-d-k'+1}{S-U+d+k'}\right)^{k'}\right] \tag{9}$$
$$\leq \binom{S}{U-d} \cdot \left[1 - \left(\frac{2D-U-d-k+1}{2U-D+k+d}\right)^{k'}\right] \tag{10}$$
$$\leq \binom{S}{U-d} \cdot \left[1 - \left(\frac{U+k+d-(2U-2D+2d+2k-1)}{U+k+d}\right)^{k'}\right] \tag{11}$$
$$\leq \binom{S}{U-d} \cdot \left[1 - \left(\frac{U+k+d-\mathcal{O}(\log n\sqrt{S})}{U+k+d}\right)^{k'}\right] \tag{12}$$
$$\leq \Theta\left(\frac{k'\log n}{\sqrt{S}}\right) \cdot \binom{S}{U-d} \tag{13}$$

□

Lemma 4. $D_{tot} \geq \binom{2S}{2D} \cdot \left[1 - \left(1 - \frac{k'}{2U+1}\right)^{k}\right]$.

change statement

Proof.

$$D_{tot} = \binom{2S}{2D} - \binom{2S}{2D-k} \tag{14}$$
$$= \binom{2S}{2D} \cdot \left[1 - \frac{(2D)(2D-1)\cdots(2D-k+1)}{(2S-2D+k)(2S-2D+k-1)\cdots(2S-2D+1)}\right] \tag{15}$$
$$\geq \binom{2S}{2D} \cdot \left[1 - \left(\frac{2D-k+1}{2S-2D+1}\right)^{k}\right] \tag{16}$$
$$\geq \binom{2S}{2D} \cdot \left[1 - \left(\frac{2U-(2U-2D+k-1)}{2U+1}\right)^{k}\right] \tag{17}$$
$$\geq \binom{2S}{2D} \cdot \left[1 - \left(\frac{(2U+1)-k'}{2U+1}\right)^{k}\right] \tag{18}$$
$$\geq \binom{2S}{2D} \cdot \left[1 - \left(1 - \frac{k'}{2U+1}\right)^{k}\right] \tag{19}$$
$$\tag{20}$$

□

Reference previous lemma

Lemma 5. *When* $kk' > 2U+1$, $D_{tot} > \frac{1}{2} \cdot \binom{2S}{2D}$.

Proof. When $kk' > 2U + 1 \implies k > \frac{2U+1}{k'}$, we will show that the above expression is greater than $\frac{1}{2}\binom{2S}{2D}$. Defining $\nu = \frac{2U+1}{k'} > 1$, we see that $(1 - \frac{1}{\nu})^k \leq (1 - \frac{1}{\nu})^\nu$. Since this is an increasing function of ν and since the limit of this function is $\frac{1}{e}$, we conclude that

$$1 - \left(1 - \frac{k'}{2U+1}\right)^k > \frac{1}{2}$$

□

Lemma 6. *When* $kk' \leq 2U+1$, $D_{tot} < c_3 \frac{k \cdot k'}{S} \cdot \binom{2S}{2D}$ *for some constant* c_3.

Proof. Now we bound the term $1 - \left(1 - \frac{k'}{2U+1}\right)^k$, given that $kk' \leq 2U + 1 \implies \frac{kk'}{2U+1} \leq 1$. Using Taylor expansion, we see that

$$\begin{aligned}
&1 - \left(1 - \frac{k'}{2U+1}\right)^k && (21)\\
&\leq \frac{kk'}{2U+1} - \frac{k(k-1)}{2} \cdot \frac{k'^2}{(2U+1)^2} && (22)\\
&\leq \frac{kk'}{2U+1} - \frac{k^2k'^2}{2(2U+1)^2} && (23)\\
&\leq \frac{kk'}{2U+1}\left(1 - \frac{kk'}{2(2U+1)}\right) && (24)\\
&\leq \frac{kk'}{2(2U+1)} \leq \frac{kk'}{\Theta(S)} && (25)\\
& && (26)
\end{aligned}$$

□