Contents

1 Presentation of the problem

1.1 Problematic

Given the state of an order book, we want to place purchase orders. Trought this document, we try to define an efficient and optimal strategy in order to reduce the cost of such orders. Since it is quite complex to define this kind of strategy, we focus on the case where we have only one purchase order to make and a time horizon before which the trade must be excuted. We have the choice between executing the order directly (market order), or placing a limit order. Since we are never sure that a limit order will be executed, we have to remove it after a certain time k and place a market order at the best offer available. Notice that the case where $k = 0$ is when a market order is executed immediately. The cost related to these kind of strategies can be of two types:

- Cost related to the spread when executing a market order.
- Cost related to the uncertainty of future price movements that we denote c.

While the first is easily measurable, the second is dependant on many factors: volatility of the market, the risk aversion of the trader... We will discuss this issue further in the following sections.

The goal of this document becomes:

> *Given the state of an order book and a single purchase order we want to make, What it is the best choice of k to minimize the cost of the trade ?*

1.2 The model and its dynamic

Let us present the basic model which our study is based on. It is the classical Cont and Larrard model. For more details please refer to their article *"Price dynamics in a Markovian limit order market"* [1]
We shall agree that most of this section is an extract from the article of Cont and Larrard. This section aims only to set up the framework of our study.

1.2.1 The model

The order book is represented by its best ask and bid queues. The state of the limit order book is represented by :

- the bid price s_t^a and the ask price s_t^b
- the size of the bid queue q_t^a representing the outstanding limit buy orders at the bid, and
- the size of the ask queue q_t^b representing the outstanding limit sell orders at the ask

The bid and ask prices are multiples of the tick size $\delta = s_t^a - s_t^b$ which we consider constant. The state of the limit order book is thus described by the triplet $X_t = (s_t^b, q_t^b, q_t^a)$ which takes values in the discrete state space $\delta\mathbb{Z} \times \mathbb{N}^2$. With no loss of generality, we can assume that :

$$\delta = 1 \tag{1}$$

We assume that orders arriving at the best bid/ask follow a poissonion process and are all of the same size 1:

- A market/cancelation order arriving at the q_t^a (resp q_t^b) reduces its size by 1 if it is not empty
- A limit order arriving at the q_t^a (resp q_t^b) adds 1 to its size

Therefore q_t^a and q_t^b follow a birth/death process.
We denote :

- $(\tau_i)_{i\geq 1}$ the sequence of random variables which stands for times between successive price changes. [1]
- $(X_i)_{i\geq 1}$ the successive moves in the price. Note that according to our hypothesis $(\delta = 1)$: $X_i \in \{-1, 1\}$.
- $(S_k)_{i\geq 0}$ the cumulative moves in the price after the n^{th} change, ie $S_n = \sum 1 \leq i \leq nX_i$. [3]

1.2.2 The dynamic

The state of the order book is modified by order book events. When the bid (resp. ask) queue is empty, the price moves down (resp. up), and (q_t^a, q_t^b) (resp.(q_t^b, q_t^a) is generated from a distribution $f(x, y)\mathrm{d}x\mathrm{d}y$ (resp. $\tilde{f}(x, y)\mathrm{d}x\mathrm{d}y$) independently from the rest of the historical variables.

This property give the market a markovian property: the new state of the market after a price move does only depend on the direction of this move. Therefore, $(X_i)_{i\geq 0}$ is markov process such as

- $\forall\, n \geq 0 \quad X_n \in \{-1, 1\}$
- $\forall\, i \geq 0 \quad \mathbb{P}(X_{i+1} = 1|X_i = 1) = \mathbb{P}(X_{i+1} = -1|X_i = -1) = p = 1 - q$

Given the symmetry of the market, we may assume that $f(x, y) = \tilde{f}(y, x)$. This assumption makes the process $(\tau_i)_i$ a sequence of IID random variables independent from $(X_i)_i$.

1.3 The optimization problem

Recall how our strategy works:

- An limit order is placed at t=0 at the best bid.
- after some time $T = \sum_{i\leq k} \tau_i$, if the order is not executed, we cancel it and make a market order at the ask side.

where k is the parameter of the strategy we seek to optimize.

Let N denote the first time where the sequence $(S_n)_{n\geq 0}$ achieves -1. Thereby :

$$N = \inf\{n \geq 0, S_n = -1\}$$

Finally, we denote c_{wait} the waiting cost per unit of time.

[1] Cont and Larrard's article presents an expression for the cumulative distribution function of each τ_i [2].
[3] see Annexe B for the generating function of S_n

The problem could be seen as a minimization problem :

$$\begin{aligned}
\min_{k\in\mathbb{N}} \quad f(k) := \min_{k\in\mathbb{N}} & \quad \mathbb{E}\left[\sum_{i=1}^{k\wedge N} \tau_i c_{wait} + (S_{k\wedge N)} + 1)\delta\right] \\
= \min_{k\in\mathbb{N}} & \quad \mathbb{E}\left[(k\wedge N)\mathbb{E}(\tau)c + (S_{k\wedge N} + 1)\right]
\end{aligned}$$

To compute the first term, we calculate the law of N analytically. For the second term, we use a Monte-Carlo Algorithm.

2 Solving the minimization problem

2.1 First time of reach for a sum of a Markov chain

In this section we adress the issue of finding the law of the variable N. Let us define the probability-generating function of N such as :

$$g_s^x(z) = \mathbb{E}(z^N|S_0 = s, X_0 = x) \quad \forall\, x \in \{-1, 1\}\ ,\ s \in \mathbb{N}\ ,\ z \in [0, 1]$$

so that

$$\mathbb{E}(z^N|S_0 = s) = \alpha g_s^1(z) + (1-\alpha) g_s^{-1}$$

where $\alpha := \mathbb{P}(X_0 = 1)$.

Given that (S_n, X_n) is a markov chain, we can infer that for all $x \in \{-1, 1\}$, $s \in \mathbb{N}$ and $z \in]0, 1[$ [4] :

$$\begin{cases} g_s^1 = z(p g_{s+1}^1 + q g_{s-1}^{-1}) & (2) \\ g_s^{-1} = z(q g_{s+1}^1 + p g_{s-1}^{-1}) & (3) \end{cases}$$

Denoting :

$$u = \frac{1}{2}\left(\frac{1}{zp} + z(2 - \frac{1}{p})\right)^{5} \tag{4}$$

and by imposing a Dirichlet boundary condition, we can prove [6] that g_s^1 dynamic is given by the following second order system :

$$\begin{cases} g_{s+1}^{-1} - 2u g_s^{-1} + g_{s-1}^{-1} = 0 & (5) \\ g_{-1}^{-1} = 1 & (6) \\ \lim\limits_{s \to +\infty} g_s^{-1} = 0 & (7) \end{cases}$$

Which is a classical linear recurrent sequence of order 2. The solution is given by :

$$g_s^{-1} = \left(u - \sqrt{u^2 - 1}\right)^{s+1} \tag{8}$$

Changing 1 to -1 in (3) we see that g_s^1 follow the same dynamic as g_s^{-1} if we change p to $1-p$, therefore we can deduce all the properties of the first from the second.

Since g_s^x is the probability-generating function of N , the law of N conditional to $S_0 = 0$ and $X_0 = x$ is given by :

$$\begin{cases} g_s^{-1}(z) = \sum\limits_{j=0}^{+\infty} \mathbb{P}_{X_0=-1}(N = j|S_0 = s) z^j & (9) \\ g_s^1(z) = \sum\limits_{j=0}^{+\infty} \mathbb{P}_{X_0=1}(N = j|S_0 = s) z^j & (10) \end{cases}$$

$\mathbb{P}_{X_0=-1}(N = j|S_0 = s)$ can be calculated using taylor expansion, and $\mathbb{P}_{X_0=1}(N = j|S_0 = s)$ can be deduced as discussed above.

[4]For clarity, we omit to note the dependence of g in z.
[5]u ≥ 1
[6]see Annexe

2.2 Esperance of stopped S_n

$E[S_{k\wedge n}]$ is calculated using Monte Carlo algorithm.

```
#include <iostream>
#include <iomanip>
#include <string>
#include <map>
#include <random>

using namespace std;

int main()
{
  int k = 5;
  double p = 0.2;

  // Number of independant simulations
  int M = 1e6;

  // generator of X_i
  random_device rd;
  mt19937 gen(rd());
  map<int, bernoulli_distribution> alea;
  // generator of X_i conditional to  X_{i-1} = 1
  alea[1] = std::bernoulli_distribution(p);
  // generator of X_i conditional to  X_{i-1} = -1
  alea[-1] = std::bernoulli_distribution(1 - p);

  double mean = 0;
  for (int m = 0; m < M; m++) {
    int S = 0;
    int X0 = 1;
    for (int i = 0; i < k && S >= 0; i++) {
      X0 = 2*alea[X0](gen)-1;
      S += X0;
    }
    mean += S;
  }

  mean /= M;
  cout << mean << endl;

return 0;
}
```

3 Numerical Results

3.1 Parmater estimation from market data

All experiments are done using the historical data of the Bund. In this section we try to estimate some parameters of the model.

$$p = \mathbb{P}(X_{i+1} = X_i) \approx \frac{\sum_{i \leq M} 1_{\{X_{i+1} = X_i\}}}{M}$$

The following progam gives $p \approx 0.2$

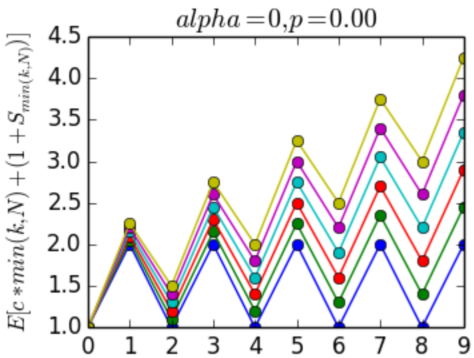
```
r=pickle.load(input)
X = r['BidPrice']
X = X[1:] - X[:-1]
X = X[ abs(X) > 1e-3] # Keep only the entries where the price change
X = np.array(map(lambda x: 1 if x > 0 else -1, X))
p = np.count_nonzero(X[1:] == X[:-1]) / len(X)
```

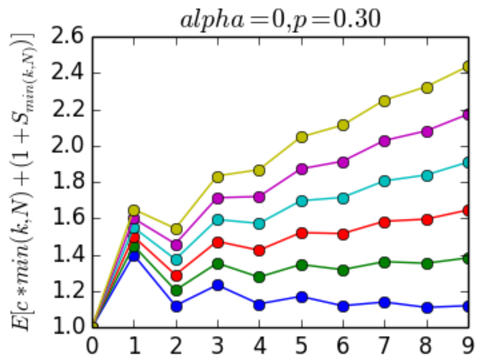
$$\mathbb{E}[\tau] \approx 3.75ms$$

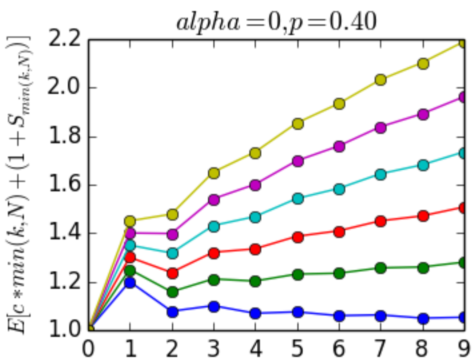
3.2 Graphical results

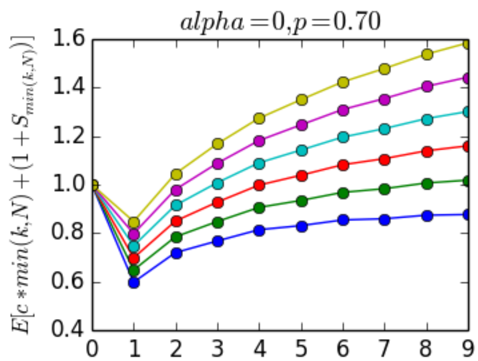
Here are the results for different combination of the parameters (p, c, α).

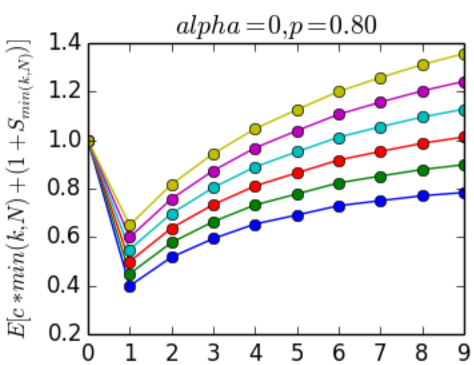
When $X_0 = -1$, ie $\alpha = 0$

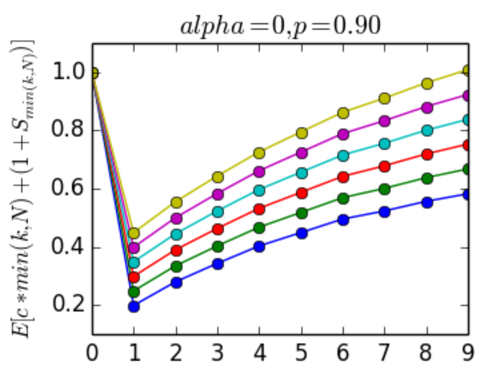


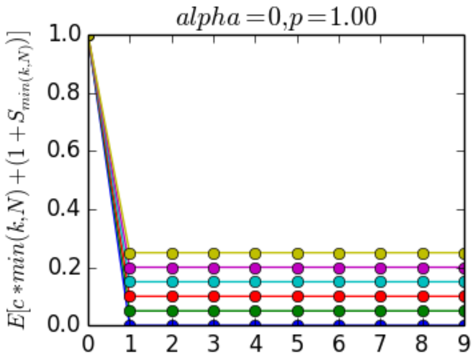






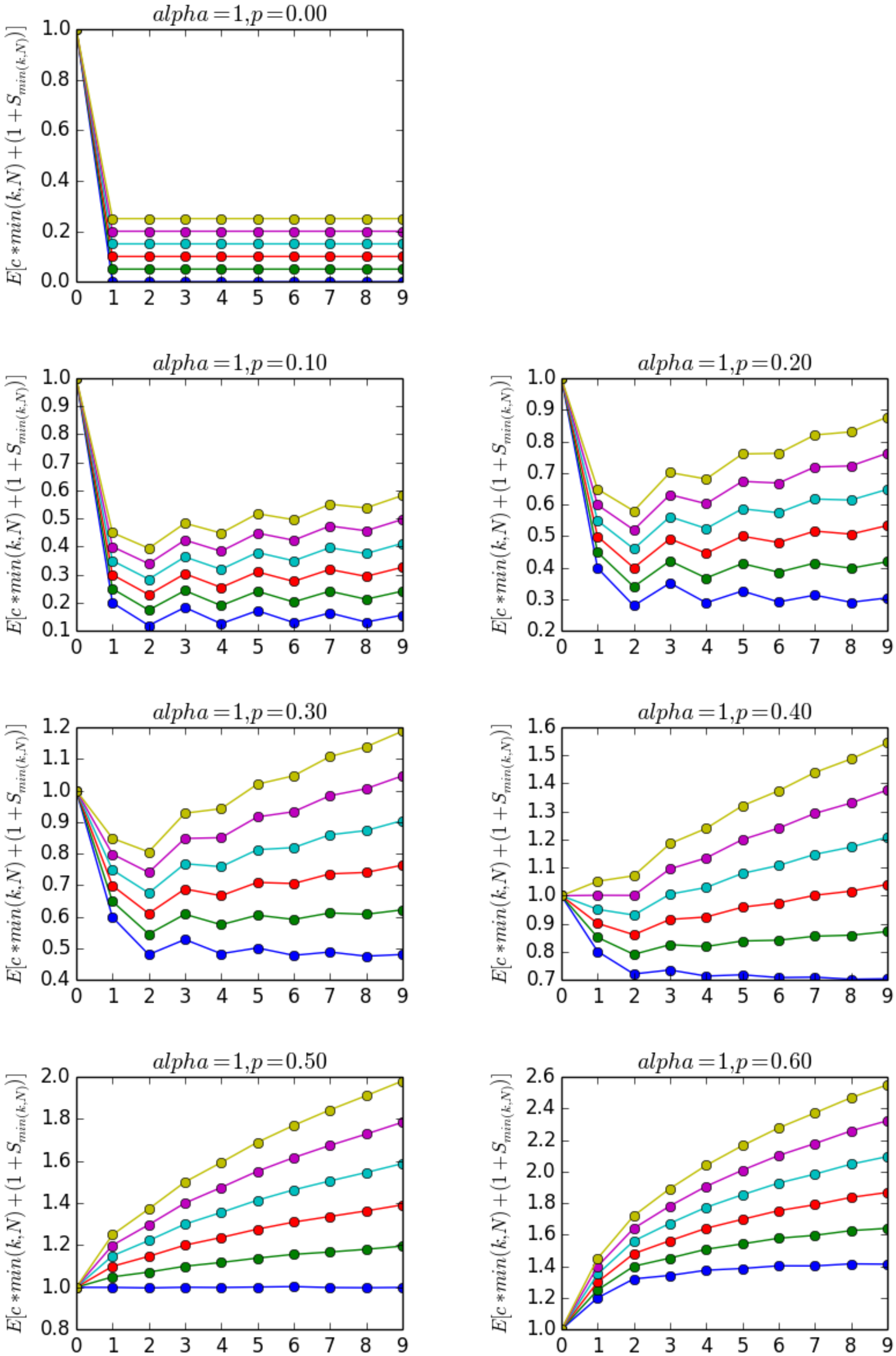


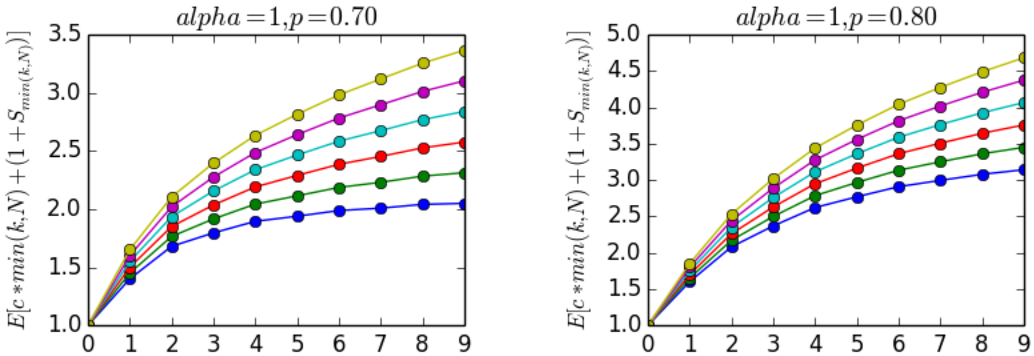




It is interesting to see that the behaviour changes when depending on wehter $p < 1/2$ or not. Indeed

When $X_0 = 1$, ie $\alpha = 1$





A Some Proofs

A.1 Reccurence equation of g_{-1}^{s}

From 3 we deduce:

$$g_{s+1}^{1} = \frac{1}{zp} g_s^{1} - \frac{q}{p} g_{s-1}^{-1}$$
$$g_{s+1}^{1} = \frac{1}{zq} g_s^{-1} - \frac{p}{q} g_{s-1}^{-1}$$

Which implies that :

$$g_s^{-1} = \frac{q}{p} g_s^{1} + \frac{z(p^2 - q^2)}{p} g_{s-1}^{-1}$$

Leading to :

$$g_{s+1}^{-1} = (\frac{1}{zp} + z(2 - \frac{1}{p})) g_s^{-1} - g_{s-1}^{-1}$$

A.2 Taylor expansion of g_{-1}^{s}

$$\begin{aligned} u &= \frac{1}{2}(\frac{1}{zp} + z(2 - \frac{1}{p}) \\ &= az + \frac{b}{z} \\ a &= 1 - \frac{1}{2p} \\ b(p) &= \frac{1}{2p} \\ a + b &= 1 \end{aligned}$$

$$
\begin{aligned}
(u^2-1)^{\frac{1}{2}} &= \sum_i \binom{\frac{1}{2}}{i}(-1)^i z^{2i-1}(b+az^2)^{1-2i} \\
&= \sum_i \binom{\frac{1}{2}}{i}(-1)^i z^{2i-1}\sum_j \binom{1-2i}{j} b^{1-2i-j}a^j z^{2j} \\
&= \sum_{i,j} \binom{\frac{1}{2}}{i}\binom{1-2i}{j}(-1)^i b^{1-2i-j}a^j z^{2(i+j)-1} \\
&= \sum_k \left(\sum_{2(i+j)-1=k} \binom{\frac{1}{2}}{i}\binom{1-2i}{j}(-1)^i b^{1-2i-j}a^j\right) z^k \\
&= \sum_r \left(\sum_{i+j=r} \binom{\frac{1}{2}}{i}\binom{1-2i}{j}(-1)^i b^{1-2i-j}a^j\right) z^{2r-1} \\
&= \sum_r \left(\sum_{i+j=r} \binom{\frac{1}{2}}{i}\binom{1-2i}{j}(-1)^i b^{1-2i-j}(1-b)^j\right) z^{2r-1} \\
&= \sum_r c_{2r-1} z^{2r-1}
\end{aligned}
$$

k impaire ≥ -1

$$c_k = \sum_{2(i+j)=k+1} \binom{\frac{1}{2}}{i}\binom{1-2i}{j}(-1)^i b^{1-2i-j}(1-b)^j$$

$$c_{-1} = \binom{\frac{1}{2}}{0}\binom{1}{0}(-1)^0 b^1(1-b)^0 = b$$

$$c_1 = \sum_{i+j=1} \binom{\frac{1}{2}}{i}\binom{1-2i}{j}(-1)^i b^{1-2i-j}(1-b)^j = a-p$$

donc

$$
\begin{aligned}
g_1^0(z) &= u-(u^2-1)^{\frac{1}{2}} \\
&= (a-c_1)z-(b+c_{-1})z^{-1}-\sum_{k>1} c_k z^k \\
&= pz-\sum_{k>1} c_k z^k \\
&= \sum_{k\geq 1} d_k z^k
\end{aligned}
$$

avec

$$d_{2r+1} = \begin{cases} p & \text{if} \quad r=0 \\ -\sum_{i\leq r+1}\binom{\frac{1}{2}}{i}\binom{1-2i}{1+r-i}(-1)^i(2p)^{2i-1}(2p-1)^{1+r-i} & \text{else} \end{cases}$$

B The law of a sum of dependent random variables

As previously, we define a Markov Chain $(X_n)_{n\geq 0}$ and its sum $(S_n)_{n\geq 0}$ In order to get the law of the r.v $(S_n)_{n\geq 0}$ we define its conditional probability-generating function as :

$$f_n^x(z) = \mathbb{E}(z^{S_n}|X_0=x) \quad \forall\ x\in\{-1,1\}\ ,\ n\in\mathbb{N}\ ,\ z\in[0,1]$$

Let us consider $x \in \{-1, 1\}$, $n \in \mathbb{N}$, $z \in]0, 1]$[7]. Thanks to the fact that (S_n, X_n) is a markov chain, we can write :

$$f_{n+1}^x(z) = z\mathbb{E}(z^{S_n}|X_0 = 1)\mathbb{P}(X_1 = 1|X_0 = x) + \frac{1}{z}\mathbb{E}(z^{S_n}|X_0 = -1)\mathbb{P}(X_1 = -1|X_0 = x) \tag{11}$$

Leading thus to the following system with the appropriate initial conditions :

$$\begin{cases} f_{n+1}^1(z) = zpf_n^1(z) + \frac{1}{z}qf_n^{-1}(z) & (12) \\ f_{n+1}^{-1}(z) = zqf_n^1(z) + \frac{1}{z}pf_n^{-1}(z) & (13) \\ f_0^1(z) = f_0^{-1}(z) = 1 & (14) \\ f_1^1(z) = zp + \frac{1}{z}q & (15) \\ f_1^{-1}(z) = zq + \frac{1}{z}p & (16) \end{cases}$$

We then prove that f_n^1 and f_n^{-1} follow the same equation, ie for $x \in \{-1, 1\}$

$$f_{n+2}^x - p(\frac{1}{z} + z)f_{n+1}^x + (2p - 1)f_n^x = 0 \tag{17}$$

Which is a classical linear recurrent sequence of order 2. The solution is given by :

$$f_n^x(z) = \frac{f_1^x - v + \sqrt{u^2 - (2p-1)}}{2\sqrt{v^2 - (2p-1)}}\Big((v + \sqrt{v^2 - (2p-1)})^n - (v - \sqrt{v^2 - (2p-1)})^n\Big) + (v - \sqrt{v^2 - (2p-1)})^n \tag{18}$$

Where :

$$v = \frac{p}{2}(z + \frac{1}{z})$$

f_n^x is the probability-generating function of $(S_n)_{n\geq 0}$. We shall notice that $\forall n \in \mathbb{N} - n \leq S_n \leq n$. We then get :

$$\begin{cases} \mathbb{E}_{X_0=x}(S_n) = \frac{\mathrm{df_n^x}}{\mathrm{d}z}(z = 1) & (19) \\ f_n^x(z) = \mathbb{E}(z^{S_n}|X_0 = x) = \sum_{-\infty}^{+\infty} \mathbb{P}(S_n = k)z^k = \sum_{-n}^{+n} \mathbb{P}(S_n = k)z^k & (20) \end{cases}$$

C Source Code

[7]The case where z=0 can be obtained by continuity.