Corrigé

1. On commence par écrire un programme Python qui donne les N premiers pas d'une marche aléatoire sur $\mathbb{Z}^d$, avec $d \in \{1, 2, 3, 4\}$.

```
from random import random
from copy import copy

def random_walk(n,d):
    if d==1:
        step = [(1,),(-1,)]
        pos = [0]
    elif d==2:
        step = [(1,0),(-1,0),(0,1),(0,-1)]
        pos = [0,0]
    elif d==3:
        step = [(1,0,0),(-1,0,0),(0,1,0),(0,-1,0),(0,0,1),(0,0,-1)]
        pos = [0,0,0]
    elif d==4:
        step = [(1,0,0,0),(-1,0,0,0),(0,1,0,0),(0,-1,0,0),
        (0,0,1,0),(0,0,-1,0),(0,0,0,1),(0,0,0,-1)]
        pos = [0,0,0,0]
    else:
        raise(NotImplementedError)
    res = []
    for k in range(n):
        st = step[int(2*d*random())]
        for i in range(d):
            pos[i] += st[i]
        res.append(copy(pos))
    return res
```

Le programme suivant calcule alors la suite $(C_n)_{1 \leq n \leq N}$:

```
def Csequence(L):
    res = []
    visited = []
    count = 0
    for x in L:
        if not (x in visited):
            count += 1
            visited.append(x)
        res.append(count)
    return res
```

Par exemple, Csequence(random_walk(30,2)) renvoie

```
[1, 2, 3, 3, 4, 5, 5, 6, 6, 6, 6, 6, 7, 8, 9, 10, 11, 11, 12, 13, 13, 14,
15, 15, 15, 16, 17, 18, 19, 20]
```

2. D'après le Théorème 1 du texte, la suite $\frac{C_n}{n}$ converge en probabilité vers α, c'est donc un estimateur consistant de ce paramètre. Les commandes Csequence(random_walk(10000,3))[−1]/10000 et Csequence(random_walk(10000,4))[−1]/10000 renvoient respectivement

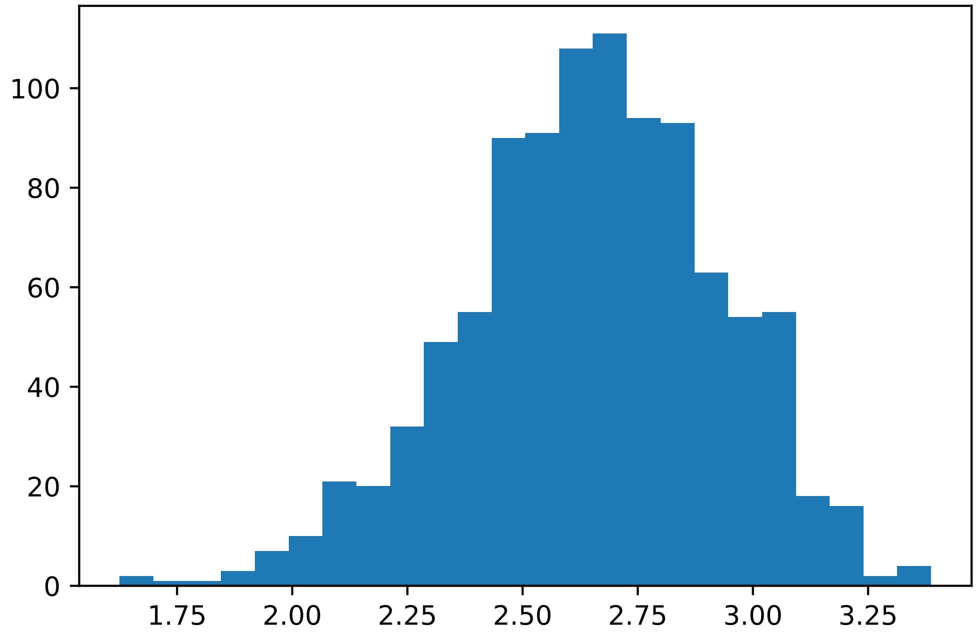
$$\alpha(d=3) \simeq 0.6731;$$
$$\alpha(d=4) \simeq 0.8172.$$

Pour $d = 3$, on est proche à 2% près du résultat théorique. Pour illustrer la convergence en probabilité de $\frac{C_n \log n}{n}$ lorsque $d = 2$, on peut dresser un histogramme de simulations des valeurs de cette variable aléatoire avec $n = 10000$. Le petit programme suivant prend quelques minutes pour dessiner un histogramme sur 1000 essais.

```
import numpy, math
import matplotlib.pyplot as plt

plt.hist(numpy.array([Csequence(random_walk(10000,2))[-1]
* math.log(10000)/10000 for i in range(1000)]), bins="auto")

plt.show()
```



L'histogramme montre bien une concentration autour d'une valeur qui semble un peu inférieure à π ; mais ceci est dû au fait qu'on a seulement une convergence vers cette valeur, et qu'une suite convergente avec des logarithmes (ici, $\frac{C_n \log n}{n}$) a souvent une vitesse de convergence très faible (c'est typiquement ce qui se produit en théorie asymptotique des nombres).

3. Si $0 \in \{X_1, \ldots, X_n\}$, alors au temps n, la marche aléatoire a visité tous les entiers compris entre m_n (négatif) et M_n (positif), donc

$$C_n = M_n - m_n + 1.$$

En fait, la relation est encore vraie si $0 \notin \{X_1, \ldots, X_n\}$: dans ce cas, la marche reste tout le temps strictement positive ou strictement négative, et si l'on regarde par exemple le cas où elle reste positive, alors $m_n = 1$, et $C_n = M_n = M_n - m_n + 1$.

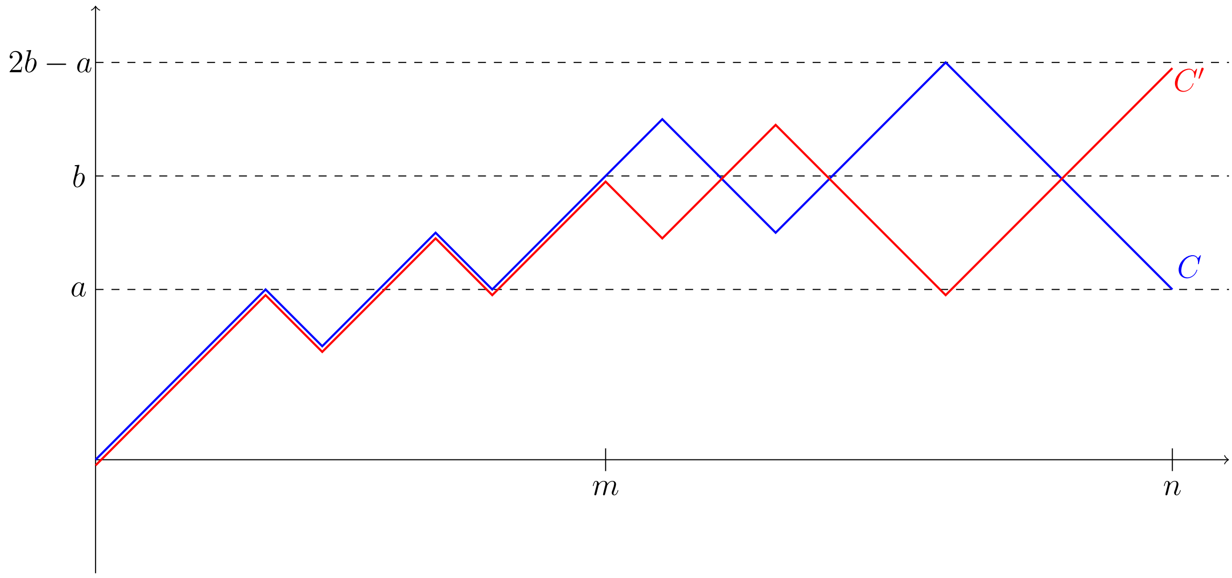
La formule ci-dessus implique que si $C_n \geq 2b+1$, alors soit $M_n \geq b$, soit $-m_n \geq b$. Par symétrie, M_n et $-m_n$ ont même loi, donc

$$\mathbb{P}[C_n \geq 2b+1] \leq \mathbb{P}[M_n \geq b] + \mathbb{P}[-m_n \geq b] = 2\,\mathbb{P}[M_n \geq b].$$

4. La première identité repose sur le principe de réflexion. Si $a \geq b$, alors il est clair que l'événement $\{M_n \geq b\}$ est inclus dans l'événement $\{X_n = a\}$, donc $\{M_n \geq b \text{ et } X_n = a\} = \{X_n = a\}$ et l'égalité des probabilités est claire. Si $a < b$, remarquons que

$$\begin{aligned}&\mathbb{P}[M_n \geq b \text{ et } X_n = a] \\ &= \frac{1}{2^n} \operatorname{card} \{C \text{ chemin de longueur } n \text{ partant de } 0, \text{ finissant en } a \text{ et dépassant le niveau } b\},\end{aligned}$$

où par chemin on entend une suite $(0 = x_0, x_1, x_2, \ldots, x_n)$ d'entiers telle que $|x_i - x_{i-1}| = 1$ pour tout $i \in [\![1, n]\!]$. En effet, tous les chemins de longueur n ont la même probabilité pour la distribution de pas $\mu = \frac{1}{2}\,\delta_1 + \frac{1}{2}\,\delta_{-1}$, à savoir $\frac{1}{2^n}$. Associons à un chemin C compté ci-dessus le chemin C' obtenu en réflechissant la partie du chemin de C après le premier temps m tel que C passe par b (voir le dessin ci-dessous).



Le chemin C' ainsi obtenu est encore de longueur n, et sa position finale est $b - (a - b) = 2b - a$. Réciproquement, tout tel chemin de longueur n s'arrêtant en $2b - a$ atteint à un temps $m \leq n$ le niveau b (car $2b - a > b$), donc est obtenu par réflexion d'un chemin C finissant en a et dépassant le niveau b. On a donc :

$$\begin{aligned}&\mathbb{P}[M_n \geq b \text{ et } X_n = a] \\ &= \frac{1}{2^n} \operatorname{card} \{C' \text{ chemin de longueur } n \text{ partant de } 0 \text{ et finissant en } 2b - a\} \\ &= \mathbb{P}[X_n = 2b - a].\end{aligned}$$

Alors, en faisant la somme sur tous les entiers a possibles, on obtient :

$$\mathbb{P}[M_n \geq b] = \sum_{a \geq b} \mathbb{P}[X_n = a] + \sum_{a < b} \mathbb{P}[X_n = 2b - a] = \mathbb{P}[X_n \geq b] + \mathbb{P}[X_n \geq b + 1].$$

En particulier, pour tout $b \geq 1$,

$$\mathbb{P}[C_n \geq 2b + 1] \leq 2\,\mathbb{P}[M_n \geq b] \leq 4\,\mathbb{P}[X_n \geq b] = 2\,\mathbb{P}[|X_n| \geq b].$$

Or, la variable X_n a moyenne 0 et variance n, donc par l'inégalité de Bienaymé–Chebyshev, $\mathbb{P}[|X_n| \geq b] \leq \frac{n}{b^2}$. Ainsi,

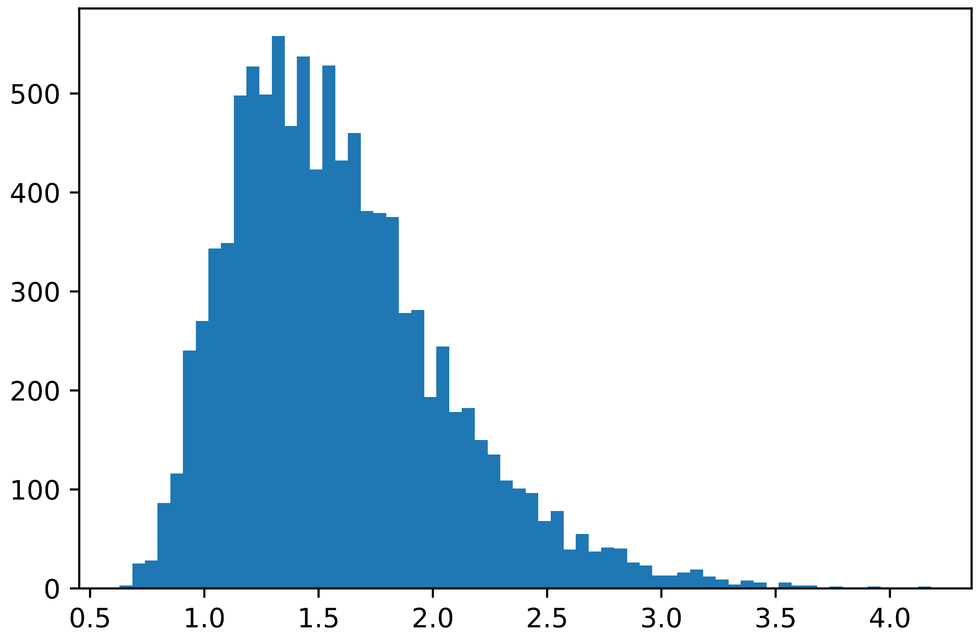
$$\mathbb{P}\left[\frac{C_n}{\sqrt{n}} \geq L\right] = O\left(\frac{1}{L^2}\right)$$

uniformément en n, et la suite de variables aléatoires $(\frac{C_n}{\sqrt{n}})_{n \geq 1}$ est tendue.

5. On dresse un histogramme de $\frac{C_n}{\sqrt{n}}$ pour $d = 1$, $n = 10000$ et $N = 10000$ essais :

```
def C1d(n):
    res = list(map(lambda x : x[0],random_walk(n,1)))
    M = max(res)
    m = min(res)
    return M-m+1

plt.hist(numpy.array([C1d(10000)/100 for i in range(10000)]), bins="auto")
plt.plot()
```



L'histogramme reste bien concentré autour de valeurs quasiment toutes plus petites que 4, et est proche d'une loi à densité sur $\mathbb{R}_+$.

6. Le texte montre que

$$\mathbb{E}[(C_n)^2] \leq \sum_{1 \leq i \leq n} \alpha_i + 2 \sum_{1 \leq i < j \leq n} \alpha_i \, \alpha_{j-i},$$

avec $\alpha_i = \mathbb{P}[R_0^+ \geq i]$. La suite $(\alpha_i)_{i \geq 1}$ est décroissante et tend vers α. Fixons $\varepsilon > 0$, et N tel que $\alpha \leq \alpha_i \leq \alpha + \varepsilon$ si $i \geq N$. La première somme $\sum_{1 \leq i \leq n} \alpha_i$ peut simplement être majorée par n, et dans la seconde somme, les termes tels que $i \geq N$ et $j \geq i + N$ donnent une contribution plus petite que $n(n-1)(\alpha + \varepsilon)^2$. Or, le nombre de couples (i, j) tels que $i < N$ ou $j < i + N$ est plus petit que $2nN$, donc

$$\mathbb{E}[(C_n)^2] \leq n(2N + 1) + n(n-1)(\alpha + \varepsilon)^2.$$

En divisant par n, on voit donc que $\mathbb{E}[(\frac{C_n}{n})^2] \leq (\alpha + \varepsilon)^2 + O(\frac{1}{n})$ pour tout $\varepsilon > 0$, puisque N est fixé (dépendant de ε). Ceci implique bien l'estimée $\alpha^2 + o(1)$.

7. Le coefficient binomial $\binom{n+m}{l}$ compte le nombre de parties A de $[\![1, n+m]\!]$ de taille l. Or, une telle partie s'écrit de manière unique sous la forme $A = A_1 \sqcup A_2$ avec $A_1 \subset [\![1, n]\!]$, $A_2 \subset [\![n+1, n+m]\!]$, $\mathrm{card}\,(A_1) = k$, $\mathrm{card}\,(A_2) = l - k$ et $0 \leq k \leq l$. Si k est fixé, alors il y a $\binom{n}{k}$ possibilités pour A_1 et $\binom{m}{l-k}$ possibilités pour A_2. Donc,

$$\binom{n+m}{l} = \sum_{k=0}^{l} \mathrm{card}\,\{(A_1, A_2)\,|\,\mathrm{card}\,(A_1) = k \text{ et } \mathrm{card}\,(A_2) = l - k\} = \sum_{k=0}^{l} \binom{n}{k}\binom{m}{l-k}.$$

8. Considérons deux marches aléatoires standards unidimensionnelles et indépendantes $(X_n)_{n\in\mathbb{N}}$ et $(X'_n)_{n\in\mathbb{N}}$. La suite $Y_n = \frac{X_n}{2}\,(1,1) + \frac{X'_n}{2}\,(1,-1)$ est une marche aléatoire standard bidimensionnelle : en effet, chacune des 4 sommes suivantes a probabilité $\frac{1}{4}$ pour un pas $Y_{n+1} - Y_n$.

$$\begin{aligned} \frac{1}{2}\,(1,1) + \frac{1}{2}\,(1,-1) &= (1,0);\\ \frac{1}{2}\,(1,1) - \frac{1}{2}\,(1,-1) &= (0,1);\\ -\frac{1}{2}\,(1,1) + \frac{1}{2}\,(1,-1) &= (0,-1);\\ -\frac{1}{2}\,(1,1) - \frac{1}{2}\,(1,-1) &= (-1,0). \end{aligned}$$

Comme $((1,1),(1,-1))$ est une base de $\mathbb{R}^2$, $Y_n = (0,0)$ si et seulement si $X_n = 0$ et $X'_n = 0$. Ainsi,

$$\mathbb{P}[Y_n = (0,0)] = \mathbb{P}[X_n = 0]\,\mathbb{P}[X'_n = 0] = (\mathbb{P}[X_n = 0])^2\,.$$

9. Pour l'identité reliant $G_{2,k}(z)$ à $G_2(z)$, notons $R_0^{(k)} = R_0^{(k)}((X_m)_{m\in\mathbb{N}})$ le temps de k-ième retour en 0 de la marche aléatoire $(X_m)_{m\in\mathbb{N}}$; comme la marche aléatoire standard bidimensionnelle est récurrente, $R_0^{(k)} < +\infty$ presque sûrement pour toute valeur de k. La variable aléatoire $R_0^{(k)}$ est un temps d'arrêt, et on notera $\mathcal{F}_{R_0^{(k)}}$ la sous-tribu de $\mathcal{F}$ qui lui est associée. Pour $k \geq 2$, remarquons que

$$R_0^{(k)}((X_m)_{m\in\mathbb{N}}) = R_0^{(k-1)}((X_m)_{m\in\mathbb{N}}) + R_0^{(1)}((X_{m+R_0^{(k-1)}})_{m\in\mathbb{N}}).$$

La propriété de Markov forte assure que la chaîne décalée en temps $(X_{m+R_0^{(k-1)}})_{m\in\mathbb{N}}$ est indépendante de $\mathcal{F}_{R_0^{(k-1)}}$ et de même loi que la chaîne de départ $(X_m)_{m\in\mathbb{N}}$ (l'indépendance vient du fait que $X_{R_0^{(k-1)}} = 0$ presque sûrement, donc la loi de la chaîne décalée conditionnellement à $\mathcal{F}_{R_0^{(k-1)}}$ est constante). Alors,

$$\begin{aligned} \mathbb{P}[R_0^{(k)}((X_m)_{m\in\mathbb{N}}) = 2n] &= \sum_{n_1+n_2=n} \mathbb{P}[R_0^{(k-1)}((X_m)_{m\in\mathbb{N}}) = 2n_1 \text{ et } R_0^{(1)}((X_{m+R_0^{(k-1)}})_{m\in\mathbb{N}}) = 2n_2]\\ &= \sum_{n_1+n_2=n} \mathbb{P}[R_0^{(k-1)}((X_m)_{m\in\mathbb{N}}) = 2n_1]\,\mathbb{P}[R_0^{(1)}((X_m)_{m\in\mathbb{N}}) = 2n_2] \end{aligned}$$

donc les coefficients de la série génératrice $G_{2,k}$ vérifient la relation de produit de Cauchy par rapport à ceux de $G_{2,k-1}$ et G_2. Par conséquent, par une récurrence immédiate, on a bien $G_{2,k}(z) = (G_2(z))^k$.

Pour la seconde identité, fixons z de module strictement plus petit que 1. On a

$$G_3(z) = \sum_{n=0}^{\infty} \mathbb{P}[R_0^+ > 2n]\,z^n,$$

la série étant absolument convergente puisque $\mathbb{P}[R_0^+ > 2n] \leq 1$. Pour tout n, $\mathbb{P}[R_0^+ > 2n] = \sum_{k>n} \mathbb{P}[R_0^+ = 2k]$, donc la famille $(\mathbb{P}[R_0^+ = 2k]\, z^n)_{(k,n)\,|\,k>n}$ est sommable. On peut donc la resommer par blocs :

$$\begin{aligned}
G_3(z) &= \sum_{(k,n)\,|\,k>n} \mathbb{P}[R_0^+ = 2k]\, z^n = \sum_{k=1}^{\infty} \mathbb{P}[R_0^+ = 2k] \left(\sum_{n=0}^{k-1} z^n \right) \\
&= \frac{1}{z-1} \left(\sum_{k=1}^{\infty} \mathbb{P}[R_0^+ = 2k]\, (z^k - 1) \right) = \frac{G_2(z) - 1}{z-1}.
\end{aligned}$$

La seconde partie de l'identité vient finalement de la Proposition 3 : $G_2(z) - 1 = -\frac{1}{G_1(z)}$.