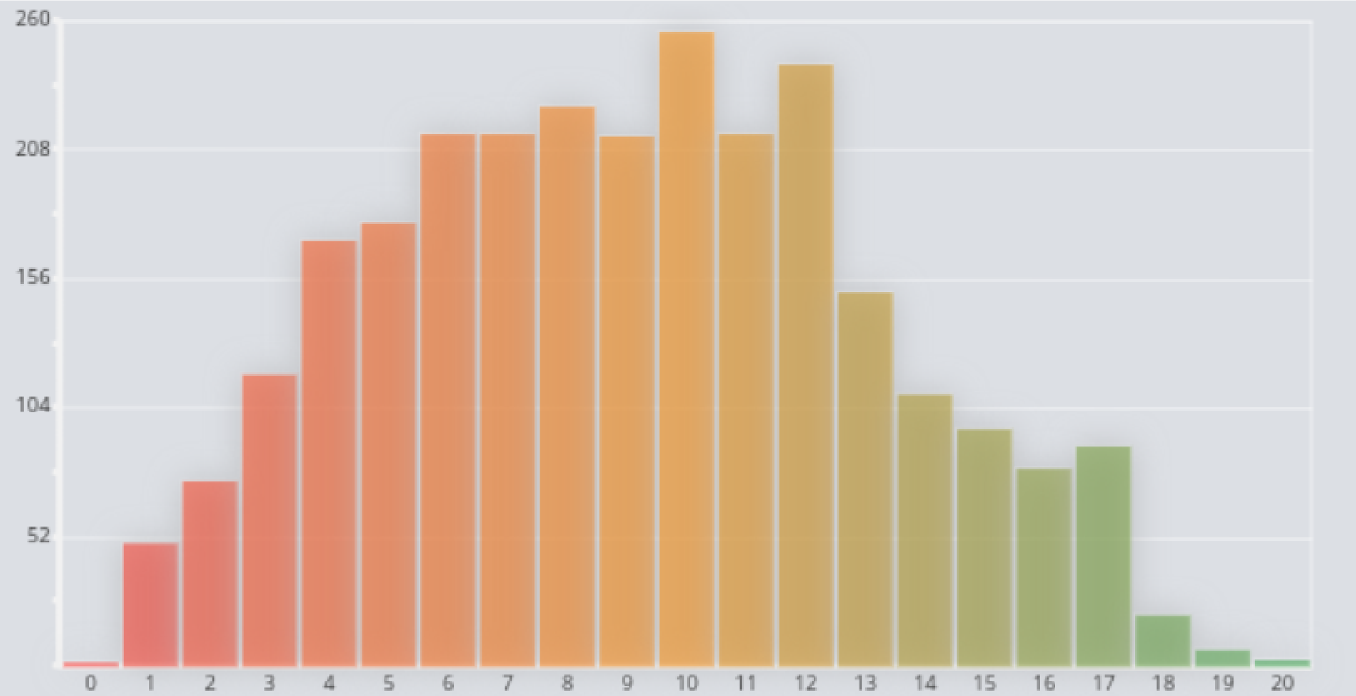
MODÉLISATION MATHÉMATIQUE ET INFORMATIQUE - 2023

1 Statistiques de l'épreuve



Minimum :	0,00 / 20	Quartile inférieur :	6,50 / 20
Moyenne :	**9,58 / 20**	**Médiane :**	**9,60 / 20**
Maximum :	20,00 / 20	Quartile supérieur :	12,50 / 20
Ecart type :	4,07	Ecart interquartile :	6,00

2 Introduction

Le sujet était un problème en 3 parties portant sur l'utilisation de marches aléatoires en modélisation.
La partie 1 traitait d'une application de la marche aléatoire classique (additive) en physique, dans le cas d'un déplacement libre (sous-partie 1.1) ou contraint dans une boîte (sous-partie 1.2). La sous partie 1.1 utilisait des résultats de probabilités (loi Binomiale, espérance, variance) et la sous partie 1.2 principalement des résultats d'algèbre linéaire (calcul matriciel, diagonalisation, produit scalaire).
Les parties 2 et 3 portaient sur l'utilisation d'une marche aléatoire multiplicative (ou modèle binomial recombinant) pour modéliser le cours d'une action (au sens de l'économie) et estimer le prix d'un contrat sur cette action appelé call. La partie 2 présentait l'objet économique et le modèle binomial recombinant de Cox-Ross-Rubinstein (CRR). La partie 2.1 testait la compréhension d'un modèle, la partie 2.2 utilisait principalement des résultats de probabilités, la partie 2.3 contenait des questions algorithmiques.
La partie 3 donnait deux façons d'estimer un paramètre du modèle de CRR appelé volatilité. Quelques questions statistiques étaient posées en partie 2.1 et la partie 3.2 contenait des questions algorithmiques autour de l'algorithme de Newton.

3 Remarques générales

Le sujet était long, la moitié des candidat· e·s n'a pas eu le temps d'investir suffisamment la dernière partie. En particulier les questions 9.(c), 11, 12, 13, 14 et 15.(a) ont été très rarement abordées. En revanche les questions 9.(b), 10.(a), 10.(b) et 15.(b) étaient accessibles et ont pu bénéficier à certains candidats lucides.

Une difficulté à cette épreuve de modélisation est d'arriver à s'extraire d'un cadre mathématique strict pour utiliser les éléments du cours dans un objectif de modélisation. La notion de contrat à terme a été bien comprise dans l'ensemble.

Les questions informatiques sont souvent bien traitées. Cette année encore, certains candidats qui n'ont pas su montrer leurs qualités mathématiques ont su en revanche déployer de réelles compétences en informatique.

La partie 3.1 proposait une ouverture sur une thématique statistique. Le jury constate que le théorème central limite est évoqué par la majorité des candidats ayant traité cette partie, bien que ses hypothèses soient peu souvent connues.

4 Difficultés mathématiques notables

- La loi binomiale et ses propriétés est bien connue dans la majorité des copies, l'indépendance est néanmoins régulièrement oubliée ;
- Le calcul de l'espérance et de la variance d'une loi est généralement bien fait ;
- Le jury constate dans certaines copies des confusions entre les variables aléatoires et les évènements ;
- Les opérations matricielles sont bien effectuées dans la majorité des copies mais les matrices et les vecteurs sont parfois inversés ;
- La diagonalisation de la question 8 a été bien traitée dans la plupart des copies ;
- Les conclusions du théorème spectral ne sont souvent pas maîtrisées.

5 Difficultés informatiques notables

- Quelques confusions dans les bornes de la fonction **range** ;
- Quelques confusions sur la différence entre **print** et **return** ;
- L'algorithme de Newton est bien maîtrisé dans la plupart des copies ;
- Quelques confusion entre le **et** et le **ou** ;
- L'utilisation de compteurs est généralement bien maîtrisée.

6 Correction partielle

6.1 Partie 1 : Marche aléatoire additive

6.1.1 Sous-Partie 1.1 : Déplacement libre

- **Question 3** Comme les sauts sont indépendants et que la probabilité d'aller à droite est de 1/2, Y_k compte le nombre de succès d'une épreuve de Bernoulli donc Y_k suit une loi binomiale $\mathcal{B}(k, 1/2)$.
- **Question 4** $X_k = X_0 + \underbrace{Y_k \times h}_{\text{à droite}} + \underbrace{(k - Y_k) \times (-h)}_{\text{à gauche}} = (2Y_k - k) \times h$
- **Question 5.(a)** L'ensemble des valeur pouvant être prises X_k est $\{(2l - k)h, l \in \{0 \ldots k\}\}$. On trouve :

$$\mathbb{P}(X_k = (k - 2l)h) = \begin{pmatrix} k \\ l \end{pmatrix} \times \left(\frac{1}{2}\right)^k$$

- **Question 5.(b)** $var(X_k) = h^2 k$

6.1.2 Sous-partie 1.2 : Déplacement contraint

- **Question 6** Soit $k \in \mathbb{N}$. On note $\overrightarrow{p}^{(k)}$ le vecteur de probabilité de X_k pour $k \in \mathbb{N}$ ce qui signifie que

$$\forall i \in [\![1, n]\!], \mathbb{P}(X_{k+1} = i) = p_i^{(k+1)}$$

Pour $i \in \{2, .., n-1\}$, l'équation revient à montrer que : $\mathbb{P}(X_{k+1} = i) = p_i^{(k+1)} = \frac{1}{2}p_{i-1}^k + \frac{1}{2}p_{i+1}^k$
Dans ce cas la particule peut venir :
- de la position $i - 1$ où elle avait une probabilité de présence p_{i-1}^k avec une probabilité $\frac{1}{2}$ de venir effectivement dans la case i
- de la position $i+1$ où elle avait une probabilité de présence $= p_{i+1}^k$ avec une probabilité $\frac{1}{2}$ de venir effectivement dans la case i

donc $\mathbb{P}(X_{k+1} = i) = p_i^{k+1} = \frac{1}{2}p_{i-1}^k + \frac{1}{2}p_{i+1}^k$.

De même pour le cas où $i = 1$ et $i = n$.
Par récurrence on montre ensuite que

$$\overrightarrow{p}^{(k)} = M^k \overrightarrow{p}^{(0)}$$

— **Question 9.(a)** La base de vecteur propre est orthogonale et l'espace propre associée à la valeur propre 1 est de dimension 1 donc pour tout $i \in [\![2, n]\!]$le produit scalaire $< \overrightarrow{v_i}, \overrightarrow{v_1} >$ est nul.

— **Question 9.(c)** Comme $\overrightarrow{v_i}$ est un vecteur propre de N associé à la valeur propre λ_i on a $N\overrightarrow{v_i} = \lambda_i \overrightarrow{v_i}$. D'où :

$$\begin{aligned} \|N\overrightarrow{v_i}\|_1 &= \|\lambda_i \overrightarrow{v_i}\|_1 \\ &= |\lambda_i|.\|\overrightarrow{v_i}\|_1 \end{aligned}$$

d'où la solution par l'inégalité (8) et le fait que $\overrightarrow{v_i}$ n'est pas nul puisqu'il s'agit d'un vecteur propre et donc sa norme 1 est strictement positive.

— **Question 10.(c)** On a $M^k = PD^kP^t$ et

$$\lim_{k \to +\infty} D^k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix}$$

donc par continuité du produit matriciel :

$$\lim_{k \to +\infty} M^k = P \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix} P^t$$

Or P est la matrice constituée des vecteurs propres orthonormés de M donc :

$$P = \begin{pmatrix} 1/\sqrt{n} & v_{2,1} & \cdots & v_{n,1} \\ \vdots & \vdots & & \vdots \\ 1/\sqrt{n} & v_{2,n} & \cdots & v_{n,n} \end{pmatrix}$$

et

$$P^t = \begin{pmatrix} 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ v_{2,1} & \cdots & v_{2,n} \\ \vdots & & \vdots \\ v_{n,1} & \cdots & v_{n,n} \end{pmatrix}$$

Ainsi :

$$P \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix} P^t = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

— **Question 11** On a $\overrightarrow{p}^{(k)} = M^k \overrightarrow{p}^{(0)}$ et

$$\lim_{k \to +\infty} M^k = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

donc par continuité du produit matriciel

$$\lim_{k \to +\infty} \overrightarrow{p}^{(k)} = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix} . \overrightarrow{p}^{(0)} = \frac{1}{n} \begin{pmatrix} \sum_{j=1}^n p_j^{(0)} \\ \vdots \\ \sum_{j=1}^n p_j^{(0)} \end{pmatrix}$$

Or comme $p^{(0)}$ est une mesure de probabilité, elle somme à 1 :

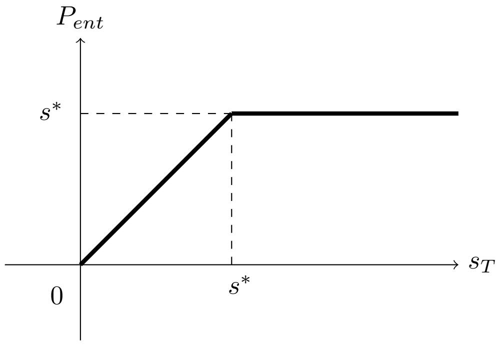
$$\lim_{k \to +\infty} \overrightarrow{p}^{(k)} = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}$$

— **Question 12** Peu importe la distribution initiale de la particule, au bout d'un temps long elle a une chance uniforme de se situer dans chaque case.

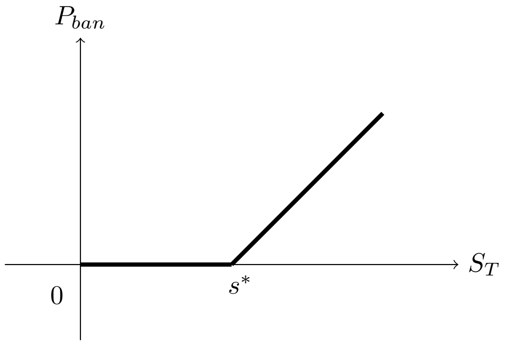
6.2 Partie 2 : Marche aléatoire multiplicative

6.2.1 Sous-partie 2.1 : Contrat à terme

- **Question 13** Plus T est grand et plus il y a d'incertitude sur ce que sera le prix du blé au temps T et notamment de risque que le prix dépasse s^* ce qui justifie que le prix du call dépende de T.
- **Question 14**
  - Si $s_T \leq s^*$, l'entreprise achète son blé au cours normal s_T d'où $P_{ent}(s_T) = s_T$
  - Si $s_T \geq s^*$, l'entreprise achète son blé au prix s^* donc $P_{ent}(s_T) = s^*$

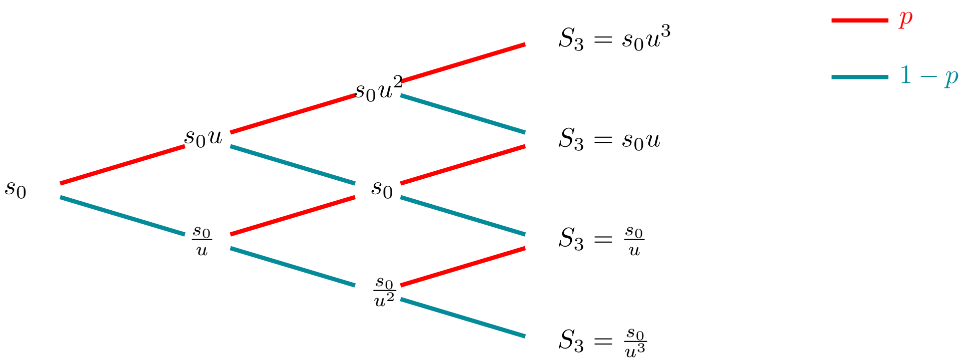


- **Question 15**
  - Si $s_T \leq s^*$, l'entreprise achète son blé au prix normal donc la banque ne verse rien d'où $P_{ban}(s_T) = 0$
  - Si $s_T \geq s^*$, l'entreprise achète son blé au prix s^* donc la banque lui rembourse la différence entre s_T et s^* donc $P_{ban}(s^T) = s^T - s^*$ (P_{ban} est une fonction affine de pente 1)



6.2.2 Sous-Partie 2.2 : Le modèle binomial recombinant de Cox-Ross-Rubinstein - application au calcul du prix du call

- **Question 1.(a)** On peut tracer l'arbre de probabilité suivant :



d'où l'ensemble $\mathcal{D}_3$. On remarque que sur cet arbre les valeurs "recombinent" grâce au fait que le coefficient de descente soit exactement l'inverse du coefficient de montée, à l'arrivée la variable S_3 ne possède donc que 4 valeurs possibles contre $8 = 2^3$ valeurs dans le cas général. Cet arbre est par conséquent l'analogue multiplicatif de celui de la partie 1.

- **Question 1.(c)** Par récurrence on peut montrer que :

$$\mathcal{D}_n = \{u^{2k-n}s_0, k \in \{0, \ldots, n\}\}$$

— **Question 2.(a)** $G_n = 0$ si $s_n \leq s^*$ et $G_n = s_n - s^*$ sinon. Comme l'ensemble des valeurs prises par S_n est $\mathcal{D}_n = \{u^{2k-n}s_0, k \in \{0,\dots,n\}\}$ par la question précédente, il s'agit donc de trouver les valeurs de $k \in \{0,\dots,n\}$ telles que $u^{2k-n}s_0 > s^*$. Or :

$$\begin{aligned} k \geq a \quad &\iff \quad k > \frac{n}{2} + \frac{\ln\left(\frac{s^*}{s_0}\right)}{2\ln(u)} \\ &\iff \quad \ln\left(u^{2k-n}\right) > \ln\left(\frac{s^*}{s_0}\right) \\ &\iff \quad u^{2k-n} > \frac{s^*}{s_0} \qquad \text{(ln est strictement croissante)} \end{aligned}$$

— **Question 2.(b)** a représente le nombre de sauts montants (les "up") nécessaires pour que l'option soit exercée c'est-à-dire pour que $s_n > s^*$.

— **Question 3.(a)** $\mathbb{E}[G_n]$ représente ce que la banque verserait en moyenne à l'entreprise si on refaisait l'expérience une infinité de fois. Si $\mathbb{E}[G_n] < c_n$ cela signifierait que la banque se verse en moyenne une somme, si $\mathbb{E}[G_n] > c_n$ qu'elle perd en moyenne de l'argent. Cette égalité signifie donc que ni la banque ni l'entreprise ne fait de gain, en moyenne.

— **Question 3.(c)** L'espérance de G_n vaut :

$$\begin{aligned} \mathbb{E}[G_n] &= \sum_{k=a}^{n} \left(u^{2k-n}s_0 - s^*\right) \mathbb{P}\left(G_n = u^{2k-n}s_0 - s^*\right) + 0 \times \mathbb{P}(G_n = 0) \\ &= \sum_{k=a}^{n} \left(u^{2k-n}s_0 - s^*\right) \times \begin{pmatrix} n \\ k \end{pmatrix} p^k(1-p)^{n-k} \end{aligned}$$

En effet la variable comptant le nombre de sauts "up" suit une loi binomiale $\mathcal{B}(n,p)$.

6.2.3 Sous-partie 2.3 : Calcul effectif du prix du call par Python

— **Question 4.(a)**
Nombre d'opérations (multiplications) : $n + k + (n-k) = 2n$
Problème : on effectue des multiplications inutiles car elles peuvent se simplifier entre numérateur et dénominateur.

— **Question 4.(b)** On simplifie les opérations entre numérateur et dénominateur.

```
def Copt(n,k):
    Num = 1
    for i in range(max([k,n-k])+1,n+1):
        Num = Num*i
    Den = 1
    for i in range(2,min([k,n-k])+1):
        Den = Den*i
    return (Num/Den)
```

— **Question 5** Cette fonction renvoie les feuilles de l'arbre de probabilité associé au modèle de Cox-Ross-Rubinstein c'est-à-dire les valeurs de l'ensemble de définition de S_n (la valeur 0 étant répétée autant de fois qu'elle apparaît dans l'arbre.

— **Question 6**

```
def call(n,u,p,S,E):
    F = Feuilles(n,u,S,E)
    c = 0
    for k in range(n+1):
        c += Copt(n,k)*F[k]*(1-p)**(n-k)*p**k
    return c
```

6.3 Partie 3 : Estimation des paramètres du modèle de Cox-Ross-Rubinstein

6.3.1 Sous-partie 3.1 : Calcul de la volatilité historique et calibrage du modèle

— **Question 7.(a)** Pour tout $i \geq 1$ la loi de $\frac{S_i}{S_{i-1}}$ ne dépend que du ième saut. Les sauts étant indépendants, les Z_i le sont également.

La loi de $Z_i = ln\left(\frac{S_i}{S_{i-1}}\right)$ est :

$\ln(u)$	$-\ln(u)$
$1/2$	$1/2$

— **Question 7.(b)**

$$\mathbb{E}[Z_1] = p \times \ln(u) + (1-p) \times (-\ln(u)) = 0$$

et

$$var(Z_1) = \mathbb{E}[Z_1^2] - \mathbb{E}[Z_1]^2 = \ln(u)^2$$

— **Question 8.(a)** Les Z_i étant indépendants et de même loi possédant une variance, le théorème central limite s'écrit :

$$\frac{\bar{Z}_n - \mathbb{E}[Z_1]}{\sqrt{var(Z_1)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

On trouve $\sigma = \sqrt{var(Z_1)} = \ln(u)$ (et on a bien $\mathbb{E}[Z_1] = 0$).

— **Question 8.(b)** Une volatilité élevée indique que le prix est instable / très variable.

— **Question 9.(a)** On a $var(\bar{Z}_n) = var(\frac{1}{n}\sum_{i=1}^n Z_i) = \frac{1}{n^2} var(\sum_{i=1}^n Z_i)$.

Les Z_i étant indépendants on a donc $var(\bar{Z}_n) = \frac{1}{n^2}\sum_{i=1}^n var(Z_i) = \frac{1}{n}\sigma^2$.

Par linéarité de l'espérance $\mathbb{E}[V_n] = \frac{1}{n-1}\sum_{i=1}^n \mathbb{E}[(Z_i - \bar{Z}_n)^2]$.

On a donc, toujours par linéarité de l'espérance

$$\begin{aligned}
\mathbb{E}[V_n] &= \frac{1}{n-1}\sum_{i=1}^n \left(\underbrace{\mathbb{E}[Z_i^2]}_{var(Z_i)} - 2\mathbb{E}[Z_i\bar{Z}_n] + \underbrace{\mathbb{E}[\bar{Z}_n^2]}_{var(\bar{Z}_n)} \right) \\
&= \frac{1}{n-1}\sum_{i=1}^n \left(\sigma^2 - 2\mathbb{E}[Z_i\bar{Z}_n] + \frac{\sigma^2}{n} \right) \\
&= \frac{(n+1)\sigma^2}{n-1} - \frac{2}{n-1}\sum_{i=1}^n \mathbb{E}[Z_i \times \frac{1}{n}\sum_{k=1}^n Z_k] \\
&= \frac{(n+1)\sigma^2}{n-1} - \frac{2}{(n-1)n}\sum_{i=1}^n \left(\underbrace{\mathbb{E}[Z_i^2]}_{=var(Z_i)=\sigma^2} + \sum_{k\neq i} \mathbb{E}[Z_i \times Z_k] \right)
\end{aligned}$$

Or par indépendance des Z_i le terme violet vaut $\frac{2}{(n-1)n}\sum_{i=1}^n \left(\sigma^2 + \sum_{k\neq i} \underbrace{\mathbb{E}[Z_i]}_{=0}\underbrace{\mathbb{E}[Z_k]}_{=0} \right)$.

Finalement

$$\mathbb{E}[V_n] = \frac{(n+1)\sigma^2}{n-1} - \frac{2}{(n-1)n} \times n\sigma^2 = \frac{n+1-2}{n-1}\sigma^2 = \sigma^2$$

— **Question 9.(b)** V_n est la variance empirique corrigée : elle représente un estimateur de σ^2 c'est-à-dire une variable aléatoire qui approche σ^2.

— **Question 9.(c)** On a

$$u = \exp(\sqrt{\sigma^2})$$

Or V_n est une *valeur approchée* du paramètre σ^2, on peut donc espérer que U_n soit une *valeur approchée* du paramètre u.

6.3.2 Sous-partie 3.2 : Calcul de la volatilité implicite avec python

— **Question 10.(c)** Si $f'(u_n) = 0$ on doit arrêter la méthode car le terme suivant de la suite n'est pas défini (division par zéro).

— **Question 11** Pour que $c_n(u) \geq 0$ il faut que $a \leq n$ c'est-à-dire que $1 + \frac{n}{2} + \frac{\ln\left(\frac{s^*}{s_0}\right)}{2\ln(u)} < n+1$ soit $\frac{\ln\left(\frac{s^*}{s_0}\right)}{\ln(u)} < n$ soit $\ln(u^n S_0) \geq \ln(s^*)$ et finalement $u \geq \left(\frac{s^*}{S_0}\right)^{\frac{1}{n}}$.

— **Question 13** Supposons n pair. Pour que a soit constant il faut que $q \leq \frac{\ln\left(\frac{s^*}{s_0}\right)}{2\ln(u)} < q+1$ avec q un entier compris entre 0 et $n-(1+\frac{n}{2})$ (et dans ce cas $a = 1+\frac{n}{2}+q$). Il faut donc :

$$2q\ln(u) \leq \ln\left(\frac{s^*}{s_0}\right) < 2(q+1)\ln(u)$$

C'est-à-dire :

$$\left(\frac{s^*}{s_0}\right)^{1/(2a-n)} < u \leq \left(\frac{s^*}{s_0}\right)^{1/(2a-n-2)}$$

De même si n est impair on retrouve le même intervalle.
Pour $a = \lfloor 1+\frac{n}{2} \rfloor$ si $u \in \left] \left(\frac{s^*}{S_0}\right)^{\frac{1}{2a-n}}, +\infty \right[$ alors s^* vérifie $u^{2a-n}S_0 \geq s^*$ et le nombre de up a est constant.

Sur ces intervalles la fonction f est dérivable, on pourra donc appliquer la méthode de Newton. Méthode proposée : on parcourt chaque intervalle pour trouver "le bon" intervalle sur lequel a est constant et u suffisamment proche du zéro de f. Comme f est croissante sur son ensemble de définition, il s'agit de trouver une valeur de a telle que $F(u(a-1)) \geq 0$ et $F(u(a)) < 0$. Une fois l'intervalle trouvé on peut appliquer la méthode de Newton car f est bien dérivable.

— **Question 15.(a)**
La première partie cherche l'intervalle sur lequel on va travailler qui vérifie $F(u(a-1)) \geq 0$ et $F(u(a)) < 0$. On initialise la recherche de Newton à $u(a)$ pour trouver u vérifiant $F(u) = 0$.

— **Question 15.(b)**

```
def volatilite(F,Fprime,s0,setoile,epsilon):
    u = mystere(F,Fprime,s0,setoile,epsilon)
    sigma = m.log(u)
    return sigma
```