Correction du sujet ESSEC / BCE 2026
Mathématiques 2 appliquées – voie générale

Sujet 287

Rectificatif. Dans la question 1, l'énoncé corrigé est bien

$$\mathbb{P}\left(\bigcup_{i=1}^{k} B_i\right) \leq \sum_{i=1}^{k} \mathbb{P}(B_i).$$

I. Résultats généraux

On rappelle que toutes les variables aléatoires sont définies sur un même espace probabilisé.

1. Inégalité de réunion. Soient $k \in \mathbb{N}^*$ et $B_1, \ldots, B_k$ des événements. Posons

$$A_1 = B_1, \qquad A_i = B_i \setminus \bigcup_{\ell=1}^{i-1} B_\ell \quad (2 \leq i \leq k).$$

Alors les A_i sont deux à deux disjoints, $A_i \subset B_i$ pour tout i, et

$$\bigcup_{i=1}^{k} B_i = \bigcup_{i=1}^{k} A_i.$$

Par additivité sur une réunion disjointe,

$$\mathbb{P}\left(\bigcup_{i=1}^{k} B_i\right) = \sum_{i=1}^{k} \mathbb{P}(A_i) \leq \sum_{i=1}^{k} \mathbb{P}(B_i).$$

C'est l'inégalité de Boole.

2. Étude de la fonction f. On fixe $\theta \in]0,1]$ et l'on pose

$$f(t) = \frac{t^2}{8} + \theta t - \ln(1 - \theta + \theta e^t), \qquad t \in \mathbb{R}.$$

a) La fonction $t \mapsto 1 - \theta + \theta e^t$ est strictement positive sur $\mathbb{R}$, donc f est bien de classe C^2 sur $\mathbb{R}$. En dérivant,

$$f'(t) = \frac{t}{4} + \theta - \frac{\theta e^t}{1 - \theta + \theta e^t},$$

puis

$$f''(t) = \frac{1}{4} - \frac{\theta(1-\theta)e^t}{(1 - \theta + \theta e^t)^2}.$$

En mettant au même dénominateur,

$$f''(t) = \frac{(1 - \theta + \theta e^t)^2 - 4\theta(1-\theta)e^t}{4(1 - \theta + \theta e^t)^2} = \frac{(1 - \theta - \theta e^t)^2}{4(1 - \theta + \theta e^t)^2}.$$

b) On a donc $f''(t) \geq 0$ pour tout t, si bien que f est convexe. De plus,

$$f(0) = 0, \qquad f'(0) = 0.$$

Le point 0 est donc un minimum global de f, d'où

$$\forall t \in \mathbb{R}, \qquad f(t) \geq 0.$$

Ceci équivaut à

$$\ln(1-\theta+\theta e^t) \leq \frac{t^2}{8} + \theta t,$$

soit encore

$$1-\theta+\theta e^t \leq \exp\left(\frac{t^2}{8} + \theta t\right).$$

En multipliant par $e^{-\theta t}$, on obtient bien

$$(1-\theta)e^{-\theta t} + \theta e^{(1-\theta)t} \leq \exp\left(\frac{t^2}{8}\right).$$

3. Convexité de l'exponentielle.

a) La fonction $t \mapsto e^t$ est convexe sur $\mathbb{R}$ car sa dérivée seconde est $e^t > 0$.

b) Soit $x \in [-\theta, 1-\theta]$. Posons

$$\lambda = \theta + x.$$

Alors $\lambda \in [0,1]$ et

$$x = \lambda(1-\theta) + (1-\lambda)(-\theta).$$

Par convexité de l'exponentielle,

$$e^{tx} \leq \lambda e^{t(1-\theta)} + (1-\lambda)e^{-\theta t}.$$

Comme $\lambda = \theta + x$ et $1-\lambda = 1-\theta-x$, on obtient

$$e^{tx} \leq (\theta+x)e^{(1-\theta)t} + (1-\theta-x)e^{-\theta t}.$$

4. Lemme exponentiel. Soit X une variable aléatoire telle que $\mathbb{E}(X) = 0$ et $X(\Omega) \subset [-\theta, 1-\theta]$. Pour tout $t \in \mathbb{R}$, la question 3.b donne presque sûrement

$$e^{tX} \leq (\theta+X)e^{(1-\theta)t} + (1-\theta-X)e^{-\theta t}.$$

En prenant l'espérance,

$$\mathbb{E}(e^{tX}) \leq \big(\theta + \mathbb{E}(X)\big)e^{(1-\theta)t} + \big(1-\theta-\mathbb{E}(X)\big)e^{-\theta t} = \theta e^{(1-\theta)t} + (1-\theta)e^{-\theta t}.$$

La question 2.b permet alors de conclure :

$$\forall t \in \mathbb{R}, \qquad \mathbb{E}(e^{tX}) \leq \exp\left(\frac{t^2}{8}\right).$$

5. Inégalité de Hoeffding. On considère $n \in \mathbb{N}^*$ et $X_1, \ldots, X_n$ indépendantes, à valeurs dans $[0,1]$, de même espérance μ. On pose

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Soit $\varepsilon > 0$.

a) Pour tout $t > 0$, l'inégalité de Markov appliquée à la variable positive $e^{t(\overline{X}_n - \mu)}$ donne

$$\mathbb{P}(\overline{X}_n - \mu \geq \varepsilon) = \mathbb{P}\big(e^{t(\overline{X}_n-\mu)} \geq e^{t\varepsilon}\big) \leq \frac{\mathbb{E}\big(e^{t(\overline{X}_n-\mu)}\big)}{e^{t\varepsilon}}.$$

Or

$$\overline{X}_n - \mu = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu),$$

donc, par indépendance,

$$\mathbb{E}\big(e^{t(\overline{X}_n-\mu)}\big) = \prod_{k=1}^{n}\mathbb{E}\left(\exp\left(\frac{t}{n}(X_k-\mu)\right)\right).$$

Ainsi

$$\mathbb{P}(\overline{X}_n - \mu \geq \varepsilon) \leq e^{-t\varepsilon}\prod_{k=1}^{n}\mathbb{E}\left(\exp\left(\frac{t}{n}(X_k-\mu)\right)\right).$$

b) Pour chaque k, la variable $X_k - \mu$ est centrée et prend ses valeurs dans $[-\mu, 1-\mu]$. La question 4 avec $\theta = \mu$ et t/n à la place de t donne

$$\mathbb{E}\left(\exp\left(\frac{t}{n}(X_k-\mu)\right)\right) \leq \exp\left(\frac{t^2}{8n^2}\right).$$

En reportant dans l'inégalité précédente,

$$\mathbb{P}(\overline{X}_n - \mu \geq \varepsilon) \leq \exp\left(-t\varepsilon + \frac{t^2}{8n}\right).$$

On choisit alors $t = 4n\varepsilon$ (minimum du trinôme en t), et l'on obtient

$$\mathbb{P}(\overline{X}_n - \mu \geq \varepsilon) \leq \exp(-2n\varepsilon^2).$$

6. Borne inférieure de Hoeffding. Appliquons la question 5 aux variables $1 - X_1, \ldots, 1 - X_n$, qui sont encore indépendantes, à valeurs dans $[0,1]$, d'espérance $1-\mu$. Leur moyenne vaut $1 - \overline{X}_n$. On en déduit

$$\mathbb{P}\big((1-\overline{X}_n) - (1-\mu) \geq \varepsilon\big) \leq e^{-2n\varepsilon^2},$$

soit

$$\mathbb{P}(\overline{X}_n - \mu \leq -\varepsilon) \leq e^{-2n\varepsilon^2}.$$

7. Intervalle de confiance. Par réunion des deux inégalités précédentes,

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq 2e^{-2n\varepsilon^2}.$$

Choisissons

$$\varepsilon = \sqrt{\frac{\ln(2/\alpha)}{2n}}.$$

Alors $2e^{-2n\varepsilon^2} = \alpha$, donc

$$\mathbb{P}\left(\mu \in \left[\overline{X}_n - \sqrt{\frac{\ln(2/\alpha)}{2n}},\ \overline{X}_n + \sqrt{\frac{\ln(2/\alpha)}{2n}}\right]\right) \geq 1 - \alpha.$$

L'intervalle demandé est donc bien un intervalle de confiance aléatoire de niveau $1-\alpha$.

II. Description du modèle, SQL et premières identités

Pour $i \in \{1, \dots, r\}$ et $j \in \{1, \dots, n\}$, on note :

$$N_{i,j} = \#\{k \in \{1, \dots, j\} \ ; \ I_k = i\}, \qquad \overline{X}_{i,j} = \frac{1}{j} \sum_{k=1}^{j} X_{i,k},$$

et

$$\widehat{X}_{i,j} = \begin{cases} \dfrac{1}{N_{i,j}} \displaystyle\sum_{k=1}^{N_{i,j}} X_{i,k} & \text{si } N_{i,j} \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

Par définition, au temps j, la récompense observée vaut

$$Y_j = X_{I_j, N_{I_j,j}}.$$

8. Requêtes SQL. Une correction possible est la suivante.

a) Pour obtenir $I_{100}(\omega)$:

```
SELECT Action
FROM Bandit
WHERE Numero = 100;
```

b) Pour obtenir la colonne des numéros des actions où l'on a joué l'action 2 et reçu une récompense non nulle :

```
SELECT Numero
FROM Bandit
WHERE Action = 2 AND Recompense <> 0;
```

c) Pour obtenir $N_{1,100}(\omega)$:

```
SELECT COUNT(*)
FROM Bandit
WHERE Numero <= 100 AND Action = 1;
```

d) Pour obtenir $\widehat{X}_{2,n}(\omega)$:

```
SELECT AVG(Recompense)
FROM Bandit
WHERE Action = 2;
```

Si $N_{2,n}(\omega) \geq 100$, alors $\widehat{X}_{2,n}(\omega)$ est une bonne estimation du paramètre p_2.

9. Espérance d'une récompense instantanée. Soit $j \in \{1, \dots, n\}$.

a) Comme $Y_j \in \{0, 1\}$, on a

$$\mathbb{E}(Y_j) = \mathbb{P}(Y_j = 1).$$

De plus, les événements $(I_j = i)_{1 \leq i \leq r}$ forment un système complet, donc

$$\mathbb{P}(Y_j = 1) = \sum_{i=1}^{r} \mathbb{P}\big((Y_j = 1) \cap (I_j = i)\big).$$

Ainsi

$$\mathbb{E}(Y_j) = \sum_{i=1}^{r} \mathbb{P}\big((Y_j = 1) \cap (I_j = i)\big).$$

b) Pour un i fixé, les événements $(N_{i,j} = k)_{1\le k\le j}$ forment une partition de $(I_j = i)$, et sur l'événement $(I_j = i) \cap (N_{i,j} = k)$ on a $Y_j = X_{i,k}$. Donc

$$\mathbb{P}\big((Y_j = 1) \cap (I_j = i)\big) = \sum_{k=1}^{j} \mathbb{P}\big((X_{i,k} = 1) \cap (I_j = i) \cap (N_{i,j} = k)\big).$$

c) La variable $X_{i,k}$ est indépendante de I_j et de $N_{i,j}$, car ces dernières dépendent seulement des choix d'actions. Donc

$$\mathbb{P}\big((X_{i,k} = 1)\cap(I_j = i)\cap(N_{i,j} = k)\big) = \mathbb{P}(X_{i,k} = 1)\,\mathbb{P}\big((I_j = i)\cap(N_{i,j} = k)\big) = p_i\,\mathbb{P}\big((I_j = i)\cap(N_{i,j} = k)\big).$$

En sommant sur k,

$$\mathbb{P}\big((Y_j = 1) \cap (I_j = i)\big) = p_i\,\mathbb{P}(I_j = i).$$

Puis en sommant sur i,

$$\mathbb{E}(Y_j) = \sum_{i=1}^{r} p_i\,\mathbb{P}(I_j = i).$$

Comme $p_i \le p^*$ pour tout i,

$$\mathbb{E}(Y_j) \le p^* \sum_{i=1}^{r} \mathbb{P}(I_j = i) = p^*.$$

10. Expression du regret moyen. On rappelle

$$R_n = \sum_{j=1}^{n} Y_j, \qquad \Delta_n = np^* - R_n, \qquad \delta_i = p^* - p_i.$$

a) Pour tout i,

$$N_{i,n} = \sum_{j=1}^{n} \mathbf{1}_{(I_j=i)}.$$

En prenant l'espérance,

$$\mathbb{E}(N_{i,n}) = \sum_{j=1}^{n} \mathbb{P}(I_j = i).$$

b) Par linéarité,

$$\mathbb{E}(\Delta_n) = np^* - \sum_{j=1}^{n} \mathbb{E}(Y_j).$$

Or la question 9.c donne

$$\mathbb{E}(Y_j) = \sum_{i=1}^{r} p_i\,\mathbb{P}(I_j = i),$$

donc

$$\mathbb{E}(\Delta_n) = \sum_{j=1}^{n}\sum_{i=1}^{r}(p^* - p_i)\mathbb{P}(I_j = i) = \sum_{i=1}^{r} \delta_i \sum_{j=1}^{n} \mathbb{P}(I_j = i).$$

En utilisant 10.a, on obtient

$$\boxed{\mathbb{E}(\Delta_n) = \sum_{i=1}^{r} \delta_i \mathbb{E}(N_{i,n}).}$$

III. Stratégie ETC et majoration du regret

11. Stratégie « No Strategy ». Dans cette question, chaque I_j suit la loi uniforme sur $\{1, \dots, r\}$. Ainsi, pour tout i,

$$\mathbb{P}(I_j = i) = \frac{1}{r}.$$

D'après la question 10.a,

$$\mathbb{E}(N_{i,n}) = \sum_{j=1}^{n} \frac{1}{r} = \frac{n}{r}.$$

Puis, grâce à 10.b,

$$\boxed{\mathbb{E}(\Delta_n) = \frac{n}{r} \sum_{i=1}^{r} \delta_i.}$$

12. Calcul de Z_2 et complétion du tableau. Ici $n = 10$, $r = 3$, $m = 2$. Pendant la phase d'exploration, les six premières actions sont

$$1, 1, 2, 2, 3, 3.$$

Les récompenses observées sont

$$Y_1 = 0,\ Y_2 = 1,\ Y_3 = 1,\ Y_4 = 1,\ Y_5 = 0,\ Y_6 = 0.$$

Donc

$$\widehat{X}_{1,6} = \frac{0+1}{2} = \frac{1}{2}, \qquad \widehat{X}_{2,6} = \frac{1+1}{2} = 1, \qquad \widehat{X}_{3,6} = \frac{0+0}{2} = 0.$$

Le maximum est atteint pour l'action 2, donc

$$Z_2 = 2.$$

La stratégie ETC joue alors toujours l'action 2 à partir de l'instant 7. La ligne $I_j(\omega)$ se complète donc par

$$I_7(\omega) = I_8(\omega) = I_9(\omega) = I_{10}(\omega) = 2.$$

13. Fonction Python calculant Z_m. Une écriture possible est :

```
def Z(m,r,Rec):
    moy = []
    for i in range(r):
        s = 0
        for k in range(i*m,(i+1)*m):
            s += Rec[k]
        moy.append(s/m)
    M = max(moy)
    for i in range(r):
        if moy[i] == M:
            return i+1
```

Cette fonction renvoie bien le plus petit indice maximisant la moyenne empirique sur la phase d'exploration.

14. Contrôle des moyennes empiriques. Soit $i \in \{1, \dots, r\}$.

a) Dans la stratégie ETC, au temps mr, chaque action a été jouée exactement m fois. Donc

$$N_{i,mr} = m.$$

Par définition,

$$\widehat{X}_{i,mr} = \frac{1}{m} \sum_{k=1}^{m} X_{i,k} = \overline{X}_{i,m}.$$

b) Les variables $X_{i,1}, \dots, X_{i,m}$ sont indépendantes, à valeurs dans $[0,1]$, d'espérance p_i. En appliquant les questions 5 et 6 à cette famille, on obtient, pour tout $\varepsilon > 0$,

$$\mathbb{P}(\widehat{X}_{i,mr} - p_i \geq \varepsilon) \leq e^{-2m\varepsilon^2}, \qquad \mathbb{P}(\widehat{X}_{i,mr} - p_i \leq -\varepsilon) \leq e^{-2m\varepsilon^2}.$$

15. Majoration de $\mathbb{E}(N_{i,n})$. Soit $s \in \{1, \dots, r\}$ tel que $p_s = p^*$. Fixons $i \in \{1, \dots, r\}$.

a) Dans ETC, on joue l'action i exactement m fois pendant l'exploration, puis encore $n - mr$ fois si et seulement si $Z_m = i$. Ainsi

$$N_{i,n} = m + (n - mr)\mathbf{1}_{(Z_m = i)}.$$

En prenant l'espérance,

$$\mathbb{E}(N_{i,n}) = m + (n - mr)\mathbb{P}(Z_m = i).$$

Si $Z_m = i$, alors $\widehat{X}_{i,mr}$ réalise le maximum des $\widehat{X}_{k,mr}$, donc en particulier

$$\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}.$$

Par conséquent,

$$\mathbb{P}(Z_m = i) \leq \mathbb{P}(\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}),$$

et donc

$$\mathbb{E}(N_{i,n}) \leq m + (n - mr)\mathbb{P}(\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}).$$

b) Supposons $\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}$. Si l'on avait simultanément

$$\widehat{X}_{i,mr} - p_i < \frac{\delta_i}{2} \qquad \text{et} \qquad p_s - \widehat{X}_{s,mr} < \frac{\delta_i}{2},$$

alors

$$\widehat{X}_{i,mr} < p_i + \frac{\delta_i}{2} = p_s - \frac{\delta_i}{2} < \widehat{X}_{s,mr},$$

ce qui est impossible. Donc

$$\{\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}\} \subset \left\{\widehat{X}_{i,mr} - p_i \geq \frac{\delta_i}{2}\right\} \cup \left\{p_s - \widehat{X}_{s,mr} \geq \frac{\delta_i}{2}\right\}.$$

Par sous-additivité,

$$\mathbb{P}(\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}) \leq \mathbb{P}\left(\widehat{X}_{i,mr} - p_i \geq \frac{\delta_i}{2}\right) + \mathbb{P}\left(p_s - \widehat{X}_{s,mr} \geq \frac{\delta_i}{2}\right).$$

c) En appliquant 14.b avec $\varepsilon = \delta_i/2$, on obtient

$$\mathbb{P}(\widehat{X}_{i,mr} \geq \widehat{X}_{s,mr}) \leq 2e^{-m\delta_i^2/2}.$$

Ainsi

$$\mathbb{E}(N_{i,n}) \leq m + 2(n - mr)e^{-m\delta_i^2/2} \leq m + 2ne^{-m\delta_i^2/2}.$$

16. Borne sur le regret ETC. Posons

$$\alpha = \sum_{i=1}^{r} \delta_i, \qquad \beta = \sum_{\substack{1 \leq i \leq r \\ \delta_i \neq 0}} \frac{1}{\delta_i}.$$

a) D'après 10.b et 15.c,

$$\mathbb{E}(\Delta_n) = \sum_{i=1}^{r} \delta_i \mathbb{E}(N_{i,n}) \leq \sum_{i=1}^{r} \delta_i \left(m + 2ne^{-m\delta_i^2/2} \right).$$

Donc

$$\boxed{\mathbb{E}(\Delta_n) \leq m \sum_{i=1}^{r} \delta_i + 2n \sum_{i=1}^{r} \delta_i e^{-m\delta_i^2/2}.}$$

b) Pour $x > 0$, la fonction $x \mapsto xe^{-x}$ est majorée par $1/e$ (maximum atteint en $x = 1$), donc

$$e^{-x} \leq \frac{1}{ex}.$$

Avec $x = m\delta_i^2/2$ (pour $\delta_i \neq 0$), on obtient

$$\delta_i e^{-m\delta_i^2/2} \leq \delta_i \frac{2}{em\delta_i^2} = \frac{2}{em\delta_i}.$$

Par suite,

$$2n \sum_{i=1}^{r} \delta_i e^{-m\delta_i^2/2} \leq \frac{4n}{em} \sum_{\delta_i \neq 0} \frac{1}{\delta_i} \leq \frac{2n\beta}{m} \qquad \left(\text{car } \frac{4}{e} < 2 \right).$$

Ainsi

$$\boxed{\mathbb{E}(\Delta_n) \leq m\alpha + \frac{2n\beta}{m}.}$$

c) Choisissons

$$m = \left\lfloor \sqrt{\frac{2n\beta}{\alpha}} \right\rfloor + 1.$$

Alors, pour n assez grand, on a bien $m \geq 2$ et $mr < n$ (car $m = O(\sqrt{n})$ alors que $n/r \to +\infty$). De plus,

$$\sqrt{\frac{2n\beta}{\alpha}} < m \leq \sqrt{\frac{2n\beta}{\alpha}} + 1.$$

Il vient donc

$$\mathbb{E}(\Delta_n) \leq \alpha \left(\sqrt{\frac{2n\beta}{\alpha}} + 1 \right) + \frac{2n\beta}{\sqrt{2n\beta/\alpha}} = \sqrt{2\alpha\beta n} + \alpha + \sqrt{2\alpha\beta n}.$$

Finalement,

$$\boxed{\mathbb{E}(\Delta_n) \leq \sqrt{8\alpha\beta}\,\sqrt{n} + \alpha.}$$

IV. Stratégie UCB et majoration logarithmique

On suppose désormais que l'on applique la stratégie UCB définie par

$$I_j = j \quad (1 \leq j \leq r),$$

et, pour $j \in \{r, \dots, n-1\}$,

$$U_{i,j} = \widehat{X}_{i,j} + \sqrt{\frac{\ln n}{N_{i,j}}}, \qquad I_{j+1} \in \arg\max_{1 \leq i \leq r} U_{i,j},$$

avec choix du plus petit indice en cas d'égalité. On note aussi

$$V_{i,j} = \overline{X}_{i,j} + \sqrt{\frac{\ln n}{j}}.$$

17. Borne élémentaire. Soient $\theta \in]0,1[$, $j \in \{1, \dots, n\}$ et $i \in \{1, \dots, r\}$. On pose

$$\gamma = \sqrt{\frac{-\ln \theta}{2j}}.$$

Alors

$$\mathbb{P}(\overline{X}_{i,j} + \gamma \leq p_i) = \mathbb{P}(\overline{X}_{i,j} - p_i \leq -\gamma).$$

La question 6 appliquée aux variables $X_{i,1}, \dots, X_{i,j}$ donne

$$\mathbb{P}(\overline{X}_{i,j} - p_i \leq -\gamma) \leq e^{-2j\gamma^2} = e^{\ln \theta} = \theta.$$

Donc

$$\boxed{\mathbb{P}(\overline{X}_{i,j} + \gamma \leq p_i) \leq \theta.}$$

18. Questions Python.

a) Au temps r, chaque action a été jouée exactement une fois, donc

$$N_{i,r} = 1 \qquad \text{et} \qquad \widehat{X}_{i,r} = X_{i,1}.$$

b) Une fonction possible est :

```
def Bernoulli1(p):
    if rd.random() < p:
        return 1
    return 0
```

c) Si `hatX` vaut $\widehat{X}_{i,j}$, `N` vaut $N_{i,j}$ et l'action choisie est i, alors après simulation d'une nouvelle récompense $Y \sim \mathcal{B}(p_i)$, la moyenne empirique devient

$$\widehat{X}_{i,j+1} = \frac{N\,\widehat{X}_{i,j} + Y}{N+1}.$$

On peut donc écrire :

```
def maj(hatX,N,i,P):
    y = Bernoulli1(P[i-1])
    return (N*hatX + y)/(N+1)
```

19. Fonction Python `Actions`. Une complétion correcte est :

```
def Actions(P,n):
    r = np.shape(P)[0]
    I = np.zeros(n, dtype=int)
    I[0:r] = [k+1 for k in range(r)]
    N = np.ones(r, dtype=int)
    hatX = np.array([Bernoulli1(P[i]) for i in range(r)], dtype=float)
    for j in range(r,n):
        Max = hatX[0] + np.sqrt(np.log(n)/N[0])
        I[j] = 1
        for k in range(1,r):
            val = hatX[k] + np.sqrt(np.log(n)/N[k])
            if val > Max:
                Max = val
                I[j] = k+1
        hatX[I[j]-1] = maj(hatX[I[j]-1], N[I[j]-1], I[j], P)
        N[I[j]-1] += 1
    return I
```

Cette fonction renvoie bien la simulation d'un vecteur des actions réalisées par l'algorithme UCB.

20. Majoration de $\mathbb{E}(N_{i,n})$. Dans toute la suite, on fixe une action optimale s telle que $p_s = p^*$ et une action sous-optimale i telle que $p_i < p^*$. Pour $u \in \{1, \dots, n-1\}$, on note

$$A_{i,u} = \left\{ \min_{j \in \{r,\dots,n-1\}} U_{s,j} \leq p_s \right\} \cup \{V_{i,u} \geq p_s\}.$$

a) Si $N_{i,n} > u$, comme la suite $(N_{i,k})_{0 \leq k \leq n}$ croît par pas de 0 ou 1, il existe $k \in \{r, \dots, n-1\}$ tel que

$$N_{i,k} = u \qquad \text{et} \qquad I_{k+1} = i.$$

Au temps $k+1$, l'algorithme choisit une action maximisant l'indice UCB, donc

$$U_{i,k} \geq U_{s,k}.$$

Or, sur l'événement $N_{i,k} = u$, on a

$$U_{i,k} = \widehat{X}_{i,k} + \sqrt{\frac{\ln n}{u}} = \overline{X}_{i,u} + \sqrt{\frac{\ln n}{u}} = V_{i,u}.$$

Par conséquent,

$$V_{i,u} \geq U_{s,k} \geq \min_{j \in \{r,\dots,n-1\}} U_{s,j}.$$

b) D'après ce qui précède,

$$\{N_{i,n} > u\} \subset A_{i,u}.$$

Comme $0 \leq N_{i,n} \leq n$,

$$\mathbb{E}(N_{i,n}) = \mathbb{E}\big(N_{i,n}\mathbf{1}_{\{N_{i,n} \leq u\}}\big) + \mathbb{E}\big(N_{i,n}\mathbf{1}_{\{N_{i,n} > u\}}\big) \leq u + n\mathbb{P}(N_{i,n} > u).$$

On en déduit

$$\boxed{\mathbb{E}(N_{i,n}) \leq u + n\mathbb{P}(A_{i,u}).}$$

21. Majoration de $\mathbb{P}(A_{i,u})$.

a) Pour tout $j \in \{r, \dots, n-1\}$, la quantité $U_{s,j}$ est de la forme $V_{s,k}$ avec $k = N_{s,j} \in \{1, \dots, n-1\}$, puisque

$$U_{s,j} = \widehat{X}_{s,j} + \sqrt{\frac{\ln n}{N_{s,j}}} = \overline{X}_{s,N_{s,j}} + \sqrt{\frac{\ln n}{N_{s,j}}} = V_{s,N_{s,j}}.$$

Ainsi

$$\left\{ \min_{j\in\{r,\dots,n-1\}} U_{s,j} \le p_s \right\} \subset \left\{ \min_{k\in\{1,\dots,n-1\}} V_{s,k} \le p_s \right\}.$$

Par l'inégalité de réunion,

$$\mathbb{P}\left(\min_{k\in\{1,\dots,n-1\}} V_{s,k} \le p_s \right) \le \sum_{k=1}^{n-1} \mathbb{P}(V_{s,k} \le p_s).$$

Or

$$V_{s,k} \le p_s \iff \overline{X}_{s,k} + \sqrt{\frac{\ln n}{k}} \le p_s,$$

et la question 17, avec $\theta = 1/n^2$, donne

$$\mathbb{P}(V_{s,k} \le p_s) \le \frac{1}{n^2}.$$

Donc

$$\mathbb{P}\left(\min_{j\in\{r,\dots,n-1\}} U_{s,j} \le p_s \right) \le \frac{n-1}{n^2} < \frac{1}{n}.$$

b) Si $\delta_i - \sqrt{\ln n/u} \ge 0$, alors

$$\{V_{i,u} \ge p_s\} = \left\{ \overline{X}_{i,u} - p_i \ge \delta_i - \sqrt{\frac{\ln n}{u}} \right\}.$$

En appliquant la question 5.b avec

$$\varepsilon = \delta_i - \sqrt{\frac{\ln n}{u}},$$

on obtient

$$\boxed{\mathbb{P}(V_{i,u} \ge p_s) \le \exp\left(-2u \left(\delta_i - \sqrt{\frac{\ln n}{u}} \right)^2 \right).}$$

c) On suppose désormais n assez grand pour que

$$\frac{4\ln n}{\delta_i^2} \in \{1, \dots, n-2\},$$

et l'on choisit

$$u = \left\lfloor \frac{4\ln n}{\delta_i^2} \right\rfloor + 1.$$

La fonction

$$\varphi(t) = \exp\Big(-2(\delta_i\sqrt{t} - \sqrt{\ln n})^2 \Big)$$

est décroissante sur $[\ln n/\delta_i^2, +\infty[$: en effet, sur cet intervalle,

$$t \longmapsto \delta_i\sqrt{t} - \sqrt{\ln n}$$

est croissante et à valeurs positives, tandis que $x \mapsto e^{-2x^2}$ est décroissante sur $[0, +\infty[$. Comme $u \geq 4\ln n/\delta_i^2$, on a

$$\delta_i\sqrt{u} - \sqrt{\ln n} \geq \sqrt{\ln n},$$

d'où

$$\mathbb{P}(V_{i,u} \geq p_s) \leq e^{-2\ln n} = \frac{1}{n^2}.$$

Enfin,

$$\mathbb{P}(A_{i,u}) \leq \mathbb{P}\left(\min_{j\in\{r,\dots,n-1\}} U_{s,j} \leq p_s\right) + \mathbb{P}(V_{i,u} \geq p_s) < \frac{1}{n} + \frac{1}{n^2} \leq \frac{2}{n}.$$

22. Borne finale sur le regret UCB. D'après 20.b,

$$\mathbb{E}(N_{i,n}) \leq u + n\mathbb{P}(A_{i,u}) \leq \left(\frac{4\ln n}{\delta_i^2} + 1\right) + 2 = \frac{4\ln n}{\delta_i^2} + 3.$$

Ainsi, pour toute action sous-optimale i,

$$\boxed{\mathbb{E}(N_{i,n}) \leq \frac{4\ln n}{\delta_i^2} + 3.}$$

En revenant à la formule du regret moyen (question 10.b),

$$\mathbb{E}(\Delta_n) = \sum_{i=1}^{r} \delta_i \mathbb{E}(N_{i,n}) \leq \sum_{\delta_i \neq 0} \delta_i \left(\frac{4\ln n}{\delta_i^2} + 3\right).$$

On obtient donc

$$\mathbb{E}(\Delta_n) \leq 4\ln n \sum_{\delta_i \neq 0} \frac{1}{\delta_i} + 3\sum_{i=1}^{r} \delta_i.$$

Avec les notations de la question 16,

$$\boxed{\mathbb{E}(\Delta_n) \leq 4\beta \ln n + 3\alpha.}$$

Fin de la correction