

Algorithme : UCB1

1. **Initialisation** : mesurer chaque bras au moins une fois
2. **Tant que** $n \prec n_{TH}$ faire :
3. Pour chaque bras i (de 1 à k) :
4. Calculer l'indice de confiance supérieur (UCB) :

$$I_i \leftarrow \hat{\mu}(i) + \sqrt{\frac{2 \ln(n)}{n_i}}$$

5. Choisir le bras $j = \arg \max(i)$ (celui avec le plus grand indice UCB)
6. Observer la récompense : $r \leftarrow \text{reward}(j)$
7. Mettre à jour la moyenne estimée du bras choisi :

$$\hat{\mu}(i) \leftarrow \hat{\mu}(i) + \frac{1}{n_j} (r - \hat{\mu}(i))$$

8. Mettre à jour le gain cumulé :

$$G_n \leftarrow G_{n-1} + r$$

9. Mettre à jour le compteur du bras choisi :

$$n_j \leftarrow n_j + 1$$

10. Incrémenter le nombre total d'itérations :

$$n \leftarrow n + 1$$

Exp3. Initialisation :

- Soit $a = A_j \in \mathcal{A}$ la stratégie choisie par le dispositif j .
- Initialiser les poids $\omega_a^j(0) = 1 \forall a \in \mathcal{A}, \forall j \in N$ et la distribution uniforme des stratégies pour chaque dispositif.
- Définir le taux d'apprentissage $y = \min\{1, \sqrt{\frac{K \log(K)}{(e-1)T}}\}$ où $e \approx 2.7182$ est la base du logarithme naturel.

Pour $t = 1$ à T faire :

1. **Initialisation** ;
2. Pour chaque dispositif terminal j faire :
 - À l'instant t , tirer une stratégie $a \in \mathcal{A}$, selon la distribution $p_a^j(t)$
 - **Si transmission** :
 - Recevoir une récompense
 - $r_a^j(t) = \begin{cases} 1 & \text{si un ACK est reçu} \\ 0 & \text{sinon} \end{cases}$
 - Mettre à jour les poids et la distribution des stratégies disponibles :
 - $\omega_a^j(t+1) = \omega_a^j \exp\left(\frac{y r_a^j(t)}{K \sum_{a=1}^K p_a^j(t)}\right)$
 - $p_a^j(t+1) = (1-y) \frac{\omega_a^j(t+1)}{\sum_{a=1}^K \omega_a^j(t+1)} + \frac{y}{K}$

○ Fin si