# Contrôle APR

M2 Informatique

16 décembre 2024

1h30

Notes de cours et TP autorisées, documentation concernant python et ses bibliothèques autorisée.

Prénom :

Nom :

Ci-dessous vous indiquez la réponse aux questions.

À l’issue du contrôle, vous m’envoyer une archive dénommée Prénom.Nom.tar.gz (vous indiquez votre nom et votre prénom) contenant :

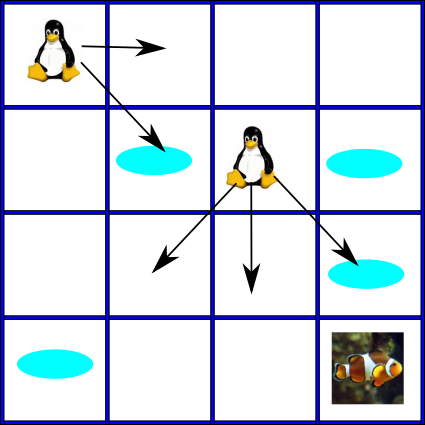
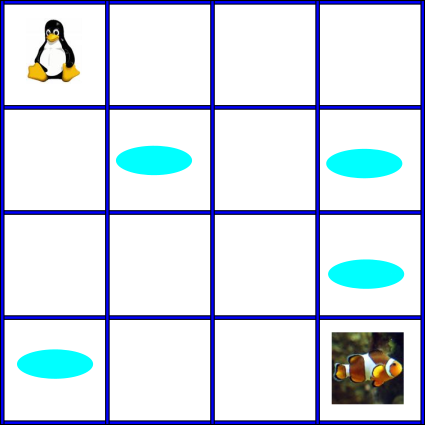
* votre compte-rendu en pdf dans un fichier dénommé Prénom.Nom.cc2.pdf,
* les programmes et une explication quant à leur utilisation pour répondre aux questions du contrôle,
* vos programmes pour les TP discrétisation et FQI.

Remarque : dans le document ci-dessous, le style des paragraphes où vous indiquez les réponses aux questions se nomme réponse. Merci d’utiliser ce style : il affiche le texte en bleu ; si votre réponse n’apparaît pas en bleu, c’est que vous n’utilisez pas le bon style.

**Conseil : réfléchissez avant d’agir !**

# Première partie sur le Q-learning tabulaire

On considère le problème du lac gelé illustré ci-dessous. Cet environnement est constitué de 16 cases. L’objectif est d’amener Tux le pingouin de son point de départ jusqu’au poisson clown, un mets extrêmement rare sur la banquise dont il raffole. Cela s’explique par le fait que Tux s’est préparé pour affronter le réchauffement climatique et la fonte de la banquise : c’est un pingouin qui ne supporte plus les eaux glaciales arctiques, préférant les eaux tropicales y vivent les poissons clown. Donc pas question pour Tux de tomber dans l’eau. Pour atteindre son mets favori, il va devoir se déplacer de case en case, une case à la fois dans l’une des quatre directions cardinales. Cependant, on sait tous que la banquise est glissante ; ainsi, quand il veut se déplacer par exemple d’une case à droite, il atteint avec équi-probabilité la case à sa droite, ou la case située au dessus ou la case située en dessous la case visée. De même, s’il fait un pas vers le bas, il atteint avec équi-probabilité la case en dessous ou celle à sa gauche ou celle à sa droite. *Etc*. Si l’une de ces cases n’existe pas, il y a équi-probabilité entre les cases qui existent. Par exemple (illustration à droite ci-dessous), de son point de départ, voulant aller une case à droite, Tux atteint avec une probabilité 0,5 la case à sa droite, ou la case en dessous ; de l’autre case, un pas vers le bas l’emmène avec une probabilité 1/3 dans l’une des 3 cases situées sous lui.



En dehors du fait que la glace est glissante, on sait aussi que marcher sur la banquise n’est pas sans risque ; si elle n’est pas assez épaisse, la glace cède sous le poids. Ainsi, les cases indiquées avec un ovale bleu indiquent des trous dans la banquise. Tux s’étant préparé pour le réchauffement climatique, il ne supporte plus l’eau glaciale du pole et il ne doit surtout pas tomber dans les trous.

Lorsque Tux atteint le poisson clown ou tombe dans un trou, l’épisode est terminé. Lorsque Tux atteint le poisson clown, Tux gagne un point ; dans tous les autres cas, Tux ne gagne rien.

Pour les actions, on prendra le codage habituel : 0 pour aller à gauche, 1 pour aller vers le haut, 2 pour aller à droite et 3 pour aller vers le bas.

À faire : résoudre ce problème par Q-Learning. Vous indiquez ci-dessous les choix que vous faites pour le Q-Learning. Vous indiquez également au bout de combien d’épisodes (moyenne, médiane, écart-type) Tux atteint le poisson clown sur 100 exécutions (si ce n’est pas trop long, sinon faites en moins et indiquez ce nombre dans votre réponse). Quand Tux atteint le poisson clown, vous indiquez le nombre de pas moyen, médian et son écart-type. Vous indiquez également la politique gloutonne trouvée par le Q-Learning dans le schéma ci-dessous à gauche et la valeur de cette politique gloutonne pour chaque état dans le schéma de droite.

**Vos choix** :

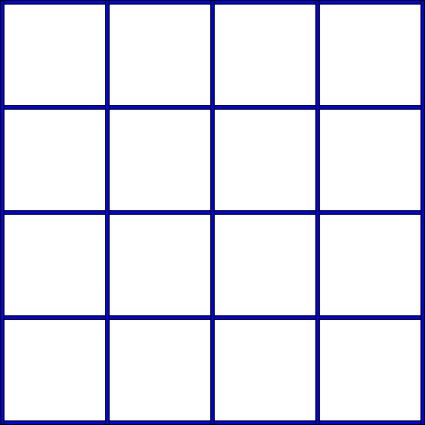
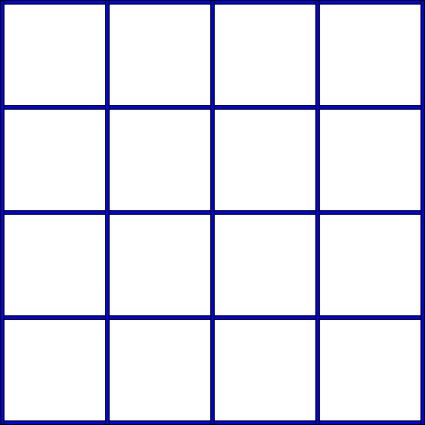
**Nombre d’exécutions réalisées** :

**Moyenne, médiane et écart-type du nombre d’épisodes pour que Tux atteigne le poisson clown** ?

**En ne considérant que les épisodes où Tux atteint le poisson clown, combien de pas fait-il en moyenne, médiane et écart-type**.

**Politique gloutonne à gauche, qualité de la l’action gloutonne dans chaque état à droite**.

Politique gloutonne Qualité de l’action gloutonne



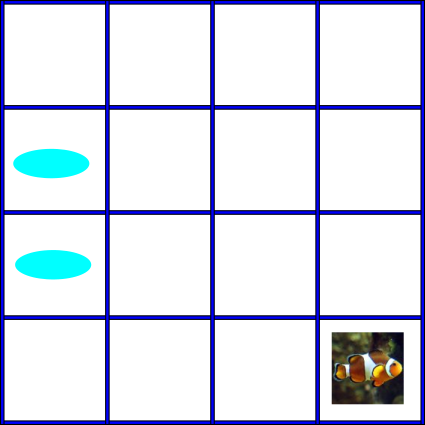
À l’issue de l’épreuve, vous m’envoyez le code avec lequel vous avez réalisé ces expérimentations et vous faites en sorte qu’en l’exécutant, j’obtienne les mêmes résultats que vous.

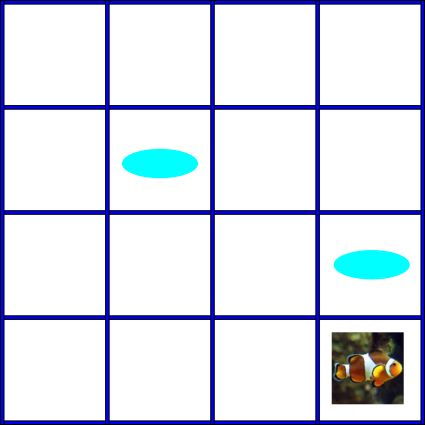
# Deuxième partie : question de réflexion

On considère à nouveau un environnement constitué de 4x4 cellules entre lesquelles on peut naviguer d’une case à la fois selon les 4 directions cardinales. Comme précédemment, l’état initial est le coin supérieur gauche et l’état d’arrivée est le coin inférieur droit. Cette fois-ci, l’environnement est déterministe : s’il fait l’action « aller à droite », Tux se déplace effectivement d’une case vers la droite ; comme d’habitude, s’il est contre un bord, celui-ci l’empêche de bouger dans cette direction (par exemple, de l’état initial, aller à gauche ou aller vers le haut laisse Tux dans la même case). Tux perçoit un retour +1 lorsqu’il atteint l’état final, 0 sinon.

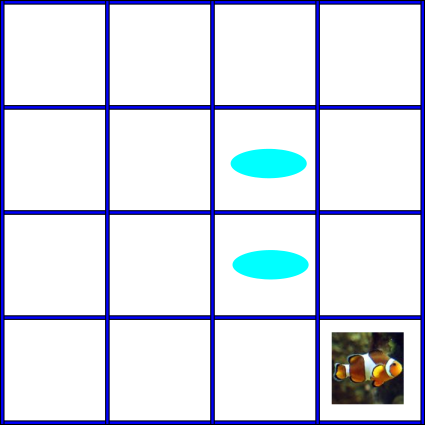
La nouveauté consiste dans le fait que des obstacles se déplacent dans l’environnement comme indiqué ci-dessous.

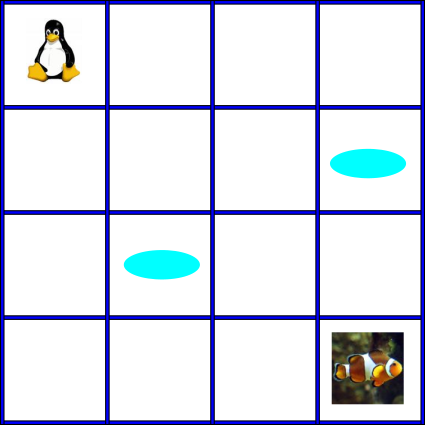
t = 0, 8, 12, ...

 t = 1, 5, 9, ...

 t = 2, 6, 10, …

t = 3, 7, 11, ...





Initialement (à gauche), les obstacles (trous dans la banquise) sont situés sous Tux. Au pas de temps suivant, l’un s’est déplacé d’une case vers la droite, l’autre d’une case vers la gauche (en supposant que l’environnement est torique pour les trous, pas pour Tux). Au pas de temps suivant, nouveau déplacement vers la droite de l’un, vers la gauche de l’autre, etc.

Ci-dessous vous indiquez en texte libre comment vous résolvez ce problème avec le Q-Learning tabulaire : est-ce que cela ne change à peu près rien par rapport à la partie 1 du contrôle ou est-ce qu’il y a une différence importante, ... ? Vous répondez aux mêmes questions que pour la partie 1. En fin d’épreuve, vous m’envoyez le code que je dois pouvoir exécuter pour reproduire vos expériences.

**Réponse**: