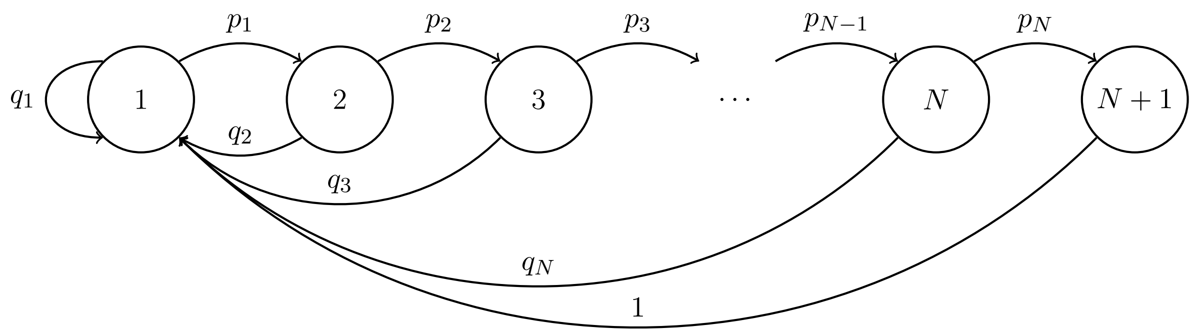
DS8 (version B)

Problème 1

Soit N un entier supérieur ou égal à 1.
On considère un joueur (ou une joueuse) qui joue à un *rogue-like*. Le rogue-like est un sous-genre de jeu vidéo dans lequel le joueur explore un donjon infesté de monstres qu'il doit combattre pour progresser vers la dernière salle, contenant le *boss final* et la promesse de récompenses épiques [(1)]. Une des caractéristiques du rogue-like est que toute mort est définitive, ce qui oblige le joueur à recommencer du début s'il souhaite continuer à jouer. Nous considérons dans la suite un joueur soumis aux règles suivantes :

- Le joueur commence au *niveau* 1 et a pour objectif d'atteindre le niveau $N+1$ afin d'y affronter le boss final et de le vaincre.
- A chaque niveau, le joueur effectue un unique *combat* qu'il peut soit gagner soit perdre.
  Plus précisément :
  - × pour tout $i \in [\![1, N]\!]$, on note $p_i \in]0, 1[$ la probabilité que le joueur gagne le combat effectué au niveau i, et on pose $q_i = 1 - p_i$. On suppose que, pour tout $i \in [\![1, N-1]\!]$, $p_i \geqslant p_{i+1}$.
  - × si le combat mené au niveau $i \in [\![1, N]\!]$ est gagné, le joueur passe au niveau supérieur, sinon, le joueur meurt et retourne au niveau 1.
  - × quelque soit l'issue du combat effectué au niveau $N+1$, le joueur recommence au niveau 1 ensuite.
- On appelle *partie* toute séquence minimale de jeu commençant au niveau 1 et se terminant soit par la mort du joueur, soit par la victoire du joueur contre le boss final.
- Le joueur enchaîne indéfiniment les parties, même en cas de victoire contre le boss final.

On représente cette chaîne de Markov à $N+1$ états par le graphe probabiliste :



On admet que toutes les v.a.r. sont définies sur un même espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. On note :

- pour tout $n \in \mathbb{N}^*$, X_n la variable aléatoire égale au numéro du niveau où se trouve le joueur lors de son n^{e} **combat**. En particulier, $X_1 = 1$.
- pour tout $j \in \mathbb{N}^*$, Y_j la variable aléatoire égale au niveau maximal atteint lors de la j^{e} **partie** jouée.
- Z la variable aléatoire égale au numéro de la **partie** où le joueur combat le boss final pour la 1^{re} fois.
- T la variable aléatoire égale au numéro du premier **combat** effectué contre le boss final.

L'objectif principal du problème est de calculer $\mathbb{E}(T)$.
Les parties **I** et **II** sont indépendantes. La partie **III** utilise les résultats de la partie **II**.

(1). Voir https://fr.wikipedia.org/wiki/Roguelike

Question préliminaire. Donner une interprétation de l'hypothèse : pour tout $i \in [\![1, N-1]\!]$, $p_i \geqslant p_{i+1}$.

Partie I : le cas $N = 1$

On pose pour simplifier $p = p_1$ et $q = 1 - p$.

1. Dessiner le graphe probabiliste dans ce contexte et avec ces notations.

2. On note, pour tout $n \in \mathbb{N}^*$, $\alpha_n = \mathbb{P}([X_n = 1])$.

a) Montrer que, pour tout $n \in \mathbb{N}^*$, $\alpha_{n+1} = 1 - p\alpha_n$.

Commentaire

On ne démontre `JAMAIS` une *relation de récurrence* par récurrence.
Il fallait ici appliquer la FPT, pour faire le lien entre la loi de X_{n+1} et de X_n. Ce sera toujours le cas lorsqu'on étudie une chaîne de Markov, il faut donc retenir cette méthode.

b) En déduire que, pour tout $n \in \mathbb{N}^*$, $\alpha_n = \dfrac{1-(-p)^n}{1+p}$.

Commentaire

Puisque la formule est donnée dans l'énoncé (pour permettre aux candidat·es de traiter la question *4* même en cas d'erreur à cette question), on peut raisonner par récurrence.
Il n'est pas faux d'appliquer la méthode du cours sur les suites arithmético-géométriques, mais on ne profite alors pas de l'énoncé. Il vaut mieux réserver cette méthode au cas où la formule n'est pas donnée, une récurrence étant plus simple à rédiger ici.

3. ***a)*** Montrer que la suite (X_n) converge en loi vers une v.a.r. X dont on explicitera la loi.

Commentaire

Donner la loi de X, c'est donner son ensemble image $X(\Omega)$ `ET` les probabilités élémentaires. Souvent, $X(\Omega)$ est oublié.

b) En déduire un état stable de la chaîne de Markov (X_n).

Commentaire

Une pure question de cours, puisque un résultat du cours nous dit que si (X_n) converge en loi vers X, alors la loi de X écrit sous forme d'un vecteur ligne π est un état stable. Il suffisait alors de recopier le résultat du calcul précédent. Le peu de succès de cette question démontre à quel point l'apprentissage du cours n'est pas fait de manière sérieuse et méthodique.

4. ***a)*** Reconnaître, pour tout $n \in \mathbb{N}^*$, la loi de $X_n - 1$.

Commentaire

Beaucoup d'erreurs sur la loi de $X_n - 1$. Oubli du cas $n = 1$ qui est particulier, confusion entre $\mathbb{P}([X_n - 1 = 0])$ et $\mathbb{P}([X_n - 1 = 1])$ ce qui amène à donner un mauvais paramètre. Une question très simple qui a donné lieu à beaucoup d'erreurs dû à un manque de rigueur.

b) Démontrer : $\forall n \in \mathbb{N}^*$, $(X_n - 1)(X_{n+1} - 1) = 0$.

Commentaire

Lorsque l'on veut démontrer une égalité entre deux variables aléatoires, il faut fixer une issue ω et faire la preuve à ω fixé. Rappellons qu'il s'agit d'une égalité entre deux **applications**.

c) En déduire : $\forall n \in \mathbb{N}^*$, $\mathrm{Cov}(X_n, X_{n+1}) = -\dfrac{(p+(-p)^n)(p+(-p)^{n+1})}{(1+p)^2}$.

Commentaire

Une question qui permettait de tester la connaissance de la bilinéarité de l'opérateur de covariance, ainsi que la capacité à faire le lien avec les questions précédentes. Très mal réussie.

d) Pour quelle(s) valeur(s) de n les variables aléatoires X_n et X_{n+1} sont indépendantes ?

Commentaire

Ici aussi, une question qui a permis de dévoiler le niveau catastrophique de maitrise du cours. La plupart des copies raisonnent comme si il y avait équivalence entre « X et Y sont indépendantes » et « $\mathrm{Cov}(X, Y) = 0$ ».

5. ***a)*** Reconnaître la loi de Z.

Commentaire

Il y a une unique manière de rédiger cette question : à l'aide des épreuves de Bernoulli et en définissant proprement le « succès ».

b) Montrer que $T = Z + 1$. En déduire $\mathbb{E}(T)$ et $\mathbb{V}(T)$.

Commentaire

Un grand classique des concours (parfois c'est $Z - 1$ au lieu de $Z + 1$). Peu d'erreurs de calculs mais encore pas mal d'oublis sur l'existence de l'espérance et de la variance de T.

Partie II : espérance conditionnelle

Soient X et Y deux variables aléatoires discrètes. On suppose que, pour tout $y \in Y(\Omega)$, $\mathbb{P}([Y = y]) \neq 0$. Pour tout événement A tel que $\mathbb{P}(A) \neq 0$, on définit, sous réserve d'existence, l'*espérance conditionnelle de X sachant A* :

$$\mathbb{E}_A(X) = \sum_{x \in X(\Omega)} x \; \mathbb{P}_A([X = x])$$

6. Soit A un événement tel que $\mathbb{P}(A) \neq 0$.

a) Démontrer que $\mathbb{P}_A$ est une application probabilité.

b) En déduire un argument non calculatoire pour justifier le fait que l'opérateur d'espérance conditionnelle sachant A est linéaire.

7. Démontrer, sous réserve d'existence et en admettant que l'on peut intervertir les sommes, la formule des espérances totales :

$$\mathbb{E}(X) = \sum_{y \in Y(\Omega)} \mathbb{P}([Y = y]) \mathbb{E}_{[Y=y]}(X)$$

Commentaire

Lorsque l'on doit démontrer une égalité entre un terme simple et un terme compliqué, et que l'on n'a pas d'idées sur la manière de procéder, il faut partir du terme compliqué et essayer de le simplifier. Le nom « formule des espérances totales » pouvait faire penser à la « formule des probabilités totales », effectivement utile ici.

Partie III : cas général

On note, pour tout $m \in [\![0, N]\!]$, $u_m = \prod_{i=1}^{m} p_i$. Par convention, $u_0 = 1$.

8. ***a)*** Justifier : $Y_1(\Omega) = [\![1, N+1]\!]$.

b) Démontrer que, pour tout $m \in [\![1, N]\!]$, $\mathbb{P}([Y_1 = m]) = u_{m-1} - u_m$.
(on pourra décomposer l'événement $[Y_1 = m]$ à l'aide des variables aléatoires X_n)

c) Calculer $\mathbb{P}([Y_1 = N+1])$.

9. ***a)*** Expliquer pourquoi les variables aléatoires Y_j suivent toutes la même loi et sont indépendantes.

b) En déduire que : $Z \hookrightarrow \mathcal{G}(u_N)$. On explicitera le calcul de $\mathbb{P}([Z = k])$ pour tout $k \in Z(\Omega)$.

10. ***a)*** Déterminer la loi conditionnelle de Y_1 sachant $[Y_1 \leqslant N]$.

Commentaire

La notion de loi conditionnelle ne semble pas connue. Elle est pourtant au programme.

b) Soient $1 \leqslant j < k$ et soit $m \in Y_j(\Omega)$. Montrer que

$$\mathbb{P}_{[Z=k]}([Y_j = m]) = \mathbb{P}_{[Y_j \leqslant N]}([Y_j = m]) = \mathbb{P}_{[Y_1 \leqslant N]}([Y_1 = m])$$

c) Soient $1 \leqslant j < k$. Montrer que

$$\mathbb{E}_{[Z=k]}(Y_j) = \frac{1}{1 - u_N} \left(\sum_{m=0}^{N} u_m - (N+1) u_N \right)$$

11. On note, pour tout $k \in \mathbb{N}$, $S_k = \sum_{j=1}^{k} Y_j$ et on pose $W = S_{Z-1}$. Autrement dit, W est la variable aléatoire définie par :

$$\forall \omega \in \Omega,\ W(\omega) = S_{Z(\omega)-1}(\omega) = \sum_{j=1}^{Z(\omega)-1} Y_j(\omega)$$

a) Démontrer : $W(\Omega) = \mathbb{N}$.

b) Soit $k \in \mathbb{N}^*$. Montrer que, pour tout $n \geqslant (k-1)N$,

$$\sum_{i=0}^{n} i\ \mathbb{P}_{[Z=k]}([S_{k-1} = i]) = \mathbb{E}_{[Z=k]}\left(S_{k-1}\right)$$

c) En déduire que, pour tout $k \in \mathbb{N}^*$, $\mathbb{E}_{[Z=k]}(W)$ existe et

$$\mathbb{E}_{[Z=k]}(W) = \mathbb{E}_{[Z=k]}\left(S_{k-1}\right)$$

12. ***a)*** Démontrer : $T = W + N + 1$.

b) En déduire que : $\mathbb{E}(T) = \dfrac{1}{u_N} \displaystyle\sum_{m=0}^{N} u_m$. Ce résultat est-il cohérent avec celui de la partie **I** ?

c) On suppose, **dans cette question uniquement**, que : pour tout $i \in [\![1, N]\!]$, $p_i = p \in]0, 1[$.

Démontrer : $\mathbb{E}(T) \underset{N\to+\infty}{\sim} \dfrac{1}{(1-p)p^N}$.

13. *Simulations informatiques.* On rappelle qu'en **Python**, si `L` est une liste, la commande `len(L)` renvoie la taille de `L` et la commande `L[-1]` permet d'accéder au dernier élément de `L`.

a) Ecrire une fonction **Python**, nommée `CalcU`, qui prend en argument une liste de probabilités `P = [`p_1`,...,`p_N`]` et qui renvoie la liste `U = [`u_0`,...,`u_N`]`.

b) Ecrire une fonction **Python**, nommée `CalcEsp`, qui prend en argument une liste de probabilités `P = [`p_1`,...,`p_N`]` et qui renvoie l'espérance de T (on utilisera la question 12b).

c) Compléter la fonction **Python** suivante, afin qu'elle prenne en argument une liste de probabilités `P = [`p_1`,...,`p_N`]` et un entier `a` spécifiant l'état de la chaîne de Markov à un instant donné et qu'elle renvoie l'état à l'instant suivant.

```
def EtapeMarkov(P,a):
    N = len(P)
    if a == _______:
        return 1
    else:
        r = rd.random()
        if r < ___________:
            return a+1
        return _____
```

Problème 2

Dans tout l'exercice, les variables aléatoires sont définies sur un même espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$.

Partie I. Préliminaires

Dans cette partie **I.**, λ désigne un réel strictement positif.

1. Soit X une variable aléatoire suivant la loi exponentielle de paramètre λ.

a) Déterminer la fonction : $x \mapsto \mathbb{P}([X > x])$ (appelée *fonction de survie de* X).

b) Pour tous nombres réels strictement positifs x et y, calculer la probabilité conditionnelle $\mathbb{P}_{[X>x]}([X > x + y])$; justifier alors que, si X modélise la durée de vie d'un phénomène, on dise de ce dernier qu'il est « sans vieillissement ».

Commentaire

L'inclusion $[X > x + y] \subset [X > x]$ a souvent été écrite à l'envers. Rappelons qu'une inclusion se traduit par une implication : $X > x + y \implies X > x$ car $y > 0$

2. Soit $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires mutuellement indépendantes, de même loi exponentielle de paramètre λ. Pour tout entier naturel n non nul, on pose : $S_n = \sum_{k=1}^{n} X_k$.

a) Déterminer l'espérance et la variance de S_n.

Commentaire

Question plutôt bien traitée dans l'ensemble, mais il y a encore des confusions entre **combinaison linéaire**, **transformation affine** et **transformation linéaire**. Prenez le temps de clarifier vos idées sur ces arguments.

b) Démontrer par récurrence que, pour tout n de $\mathbb{N}^*$, la variable aléatoire S_n admet pour densité la fonction :

$$f_n : t \mapsto \begin{cases} 0 & \text{si } t < 0\,; \\ \dfrac{\lambda^n}{(n-1)!}\, \mathrm{e}^{-\lambda t}\, t^{n-1} & \text{si } t \geqslant 0. \end{cases}$$

Pour cela, on admettra que, si U et V sont des variables aléatoires indépendantes admettant respectivement pour densité les fonctions f_U et f_V, alors la variable aléatoire $U + V$ admet pour densité la fonction f_{U+V} définie sur $\mathbb{R}$ par :

$$f_{U+V}(t) = \int_{-\infty}^{+\infty} f_U(x)\ f_V(t-x)\ dx$$

Commentaire

Il faut penser à vérifier les hypothèses du théorème donné par l'énoncé avant de l'appliquer.

II. Loi de Pareto (Vilfredo Pareto (1848-1923), sociologue et économiste italien)

Soient a et b des réels strictement positifs. Par définition, on dit d'une variable aléatoire qu'elle suit la loi de Pareto de paramètres a et b si elle admet pour densité la fonction f définie sur $\mathbb{R}$ par :

$$\forall t \in \mathbb{R},\ f(t) = \begin{cases} 0 & \text{si } t < b\,; \\ a\dfrac{b^a}{t^{a+1}} & \text{si } t \geqslant b. \end{cases}$$

Soit alors X une variable aléatoire de loi de Pareto de paramètres a et b.

3. Vérifier que l'égalité : $\displaystyle\int_{-\infty}^{+\infty} f(t)\ dt = 1$ est bien satisfaite ; calculer l'espérance et la variance de X, en précisant à quelles conditions chacune de ces quantités existe.

> **Commentaire**
>
> Bien que l'énoncé demande explicitement des conditions d'existence, assez peu de copies s'embarrassent avec la vérification de l'existence des objets manipulés. Les conditions d'existence sont donc passées sous silence, comme si l'existence *allait de soi*. Ce n'est pas le cas.

4. Déterminer la fonction de répartition de X. Préciser la fonction de survie : $x \mapsto \mathbb{P}([X > x])$.

5. Démontrer que, pour tout réel y positif ou nul, la probabilité conditionnelle $\mathbb{P}_{[X>x]}([X > x + y])$ tend vers 1 quand x tend vers $+\infty$. De façon analogue à la question **I.1.b)**, que peut-on dire d'un phénomène dont la durée de vie est modélisée par X ?

6. On pose dans cette question : $Y = \ln\left(\dfrac{X}{b}\right)$.

a) Démontrer que Y suit une loi exponentielle dont on précisera le paramètre.

> **Commentaire**
>
> La croissance de l'exponentielle est souvent citée, mais sans le caractère strict. Or, on a besoin de la **stricte** croissance dès que l'on raisonne par équivalence. L'égalité entre deux événements est un raisonnement par équivalence (double inclusion des ensembles).

b) Déduire de la question précédente une fonction **Python**, nommée `SimulX(a,b)`, qui permette de simuler la variable aléatoire X.

III. Estimation des paramètres d'une loi de Pareto

Les instants aléatoires des arrivées de paquets (symboles binaires représentant de l'information de type audio, vidéo, données, ...) dans un canal de communication sont modélisés par une variable aléatoire X suivant une loi de Pareto de paramètres α et β ($\alpha > 0$, $\beta > 0$).
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, toutes de même loi que X.
On dit qu'un estimateur T_n de θ est *sans biais* si $\mathbb{E}(T_n) = \theta$.

7. On suppose tout d'abord que le paramètre β fait partie des caractéristiques connues du canal de communication ; on se propose de déterminer un estimateur de α par une méthode dite « du maximum de vraisemblance ». Pour cela, n désignant un entier naturel non nul et $x_1, \ldots, x_n$ des réels supérieurs ou égaux à β, on introduit la fonction L, à valeurs dans $\mathbb{R}$ et définie sur $]0, +\infty[$ par :

$$L(a) = f_a(x_1) \times \cdots \times f_a(x_n) \ = \ \prod_{k=1}^{n} f_a\left(x_k\right),$$

où f_a est la fonction définie sur $\mathbb{R}$ par :

$$\forall t \in \mathbb{R},\ f_a(t) = \begin{cases} 0 & \text{si } t < \beta\,; \\ a\dfrac{\beta^a}{t^{a+1}} & \text{si } t \geqslant \beta. \end{cases}$$

a) Exprimer $L(a)$, puis $\ln\left(L(a)\right)$.

b) On considère la fonction φ de $]0,+\infty[$ dans $\mathbb{R}$:

$$a \mapsto n\ln(a) + na\ln(\beta) - (a+1)\sum_{k=1}^{n} \ln(x_k)$$

i. Démontrer que la fonction φ admet un maximum, atteint en un unique réel w que l'on exprimera en fonction de x_1, ..., x_n et de β.

ii. Que peut-on dire de w pour la fonction L ?

c) On pose dorénavant, pour tout n de $\mathbb{N}^*$:

$$W_n = \frac{n}{\sum_{k=1}^{n} \ln\left(\frac{X_k}{\beta}\right)}$$

(la suite $(W_n)_{n\geqslant 1}$ est appelée *estimateur du maximum de vraisemblance*)

i. Justifier que la variable aléatoire $\sum_{k=1}^{n} \ln\left(\frac{X_k}{\beta}\right)$ admet pour densité la fonction f_n définie dans **I.2.b)** en prenant $\lambda = \alpha$.

ii. À l'aide du théorème de transfert, en déduire que W_n admet pour espérance $\frac{n\alpha}{n-1}$ lorsque $n \geqslant 2$, puis proposer un estimateur sans biais de α construit sur W_n.

d) On pose, pour tout n de $\mathbb{N}^*$: $W'_n = \dfrac{n-1}{n}\, W_n$.

i. Soit n un entier supérieur ou égal à 3.

En admettant que le moment d'ordre 2 de W'_n est égal à $\dfrac{(n-1)\,\alpha^2}{n-2}$, calculer la variance de W'_n puis établir, à l'aide de l'inégalité de Bienaymé-Tchebychev, que pour tout réel ε strictement positif, on a l'inégalité :

$$\mathbb{P}\left(\left[W'_n - \varepsilon < \alpha < W'_n + \varepsilon\right]\right) \geqslant 1 - \frac{\alpha^2}{\varepsilon^2\,(n-2)}$$

ii. On suppose dans cette question (et elle seule) que α est strictement compris entre 1 et 2.

Déterminer un entier naturel N tel que, pour tout entier n supérieur ou égal à N, $\left[W'_n - \dfrac{1}{10}, W'_n + \dfrac{1}{10}\right]$ soit un intervalle de confiance du paramètre α au niveau de confiance 0,95.

8. On suppose maintenant que seul le paramètre α est déjà identifié et qu'il vérifie : $\alpha > 2$.

a) Pour tout entier strictement positif n, on pose : $Y_n = c_n \sum\limits_{k=1}^{n} X_k$, où le réel c_n est choisi de sorte que Y_n soit un estimateur sans biais de β.

i. Calculer c_n.

ii. Quelle est la limite de la variance de Y_n quand n tend vers $+\infty$? (on dit que l'estimateur est convergent)

b) Pour tout entier strictement positif n, on pose : $Z_n = \min(X_1, \dots, X_n)$.

i. Déterminer la fonction de répartition de Z_n, puis reconnaître sa loi et préciser son espérance. Quelle est la limite de cette dernière quand n tend vers $+\infty$?

ii. Pour tout entier strictement positif n, on pose : $Z'_n = d_n\, Z_n$, où le réel d_n est choisi de telle sorte que $(Z'_n)_{n\geqslant 1}$ soit un estimateur sans biais de β.

Quelle est la limite de la variance de Z'_n quand n tend vers $+\infty$?

iii. Démontrer que l'estimateur $(Z'_n)_{n\geqslant 1}$ est plus efficace que l'estimateur $(Y_n)_{n\geqslant 1}$, c'est-à-dire, qu'à partir d'un certain rang, la variance de Z'_n est inférieure à celle de Y_n.