CAPES Maths 2021, épreuve 1, parties D et E : Markov, Bienaymé-Tchebychev Le problème du collectionneur

De cette épreuve 1, je ne propose des corrigés que des parties D et E.

Les parties A, B et C ont toutes été évoquées dans des problèmes antérieurs. Il en est ainsi par exemple du « problème de Bâle » qui fait l'objet d'un sujet tiré d'un ancien problème de Baccalauréat série C.

Comme chaque année depuis 2018, au moins 50 % des thèmes abordés dans les sujets des écrits du CAPES font l'objet d'une étude figurant quelque part dans les pages d'Ecrit. Encore faut-il trouver les « bonnes » études, j'en conviens. Le sujet 2021 ne déroge pas à la règle.

Partie D : Inégalités de Markov et de Bienaymé-Tchebychev.

XIX. Inégalité de Markov

La question n'indique pas la nature de la variable aléatoire Y, elle peut être finie ou discrète ou bien à densité. La démonstration n'est pas la même suivant le cas. Nous nous en tenons à des variables finies ou discrètes, compte tenu de l'ambiance du problème et de l'application qui en sera donnée dans la partie E.

- Soit Y une variable aléatoire finie ou discrète positive définie sur un univers Ω, et possédant une espérance. Elle en possède certainement une si elle est finie, et, si elle est discrète, la série $E(Y)=\sum_{k=1}^{+\infty} y_k \times P(Y=y_k)$ est une série à termes positifs (car Y est par hypothèse à valeurs positives) convergente. Considérons que les variables aléatoires finies se mettent dans le sac des discrètes, elles ont un nombre fini de termes non nuls.
- Soit a un réel strictement positif.

Comme indiqué dans l'énoncé de cette question, décomposons l'ensemble $Y(\Omega)$ des valeurs prises par Y en deux sous ensembles disjoints : $Y_1=[Y\geq a]=\{y\in Y(\Omega)\ ;\ y\geq a\}$ et $Y_2=[Y<a]=\{y\in Y(\Omega)\ ;\ y<a\}$, ces deux sous ensembles formant une partition de $Y(\Omega)$.

- Puisque la série $E(Y)=\sum_{k=1}^{+\infty} y_k \times P(Y=y_k)$ est une série à termes positifs convergente, sa somme ne dépend pas de l'ordre de sommation. Il est légitime de changer l'ordre de sommation, cela ne modifie pas la valeur de la somme : $E(Y)=\sum_{y_k\in Y_1} y_k \times P(Y=y_k)+\sum_{y_k\in Y_2} y_k \times P(Y=y_k)$

- Puisque Y est une variable aléatoire positive, $y_k \times P(Y = y_k)$ est positif ou nul pour tout k et la somme $\sum_{y_k \in Y_2} y_k \times P(Y = y_k)$ est positive ou nulle. Si on néglige cette somme dans l'expression de l'espérance, on obtient une inégalité : $E(Y) \geq \sum_{y_k \in Y_1} y_k \times P(Y = y_k)$
- Mais lorsque $y_k \in Y_1$, l'inégalité $y_k \geq a$ est vérifiée. Donc : $E(Y) \geq \sum_{y_k \in Y_1} a \times P(Y = y_k) = a \sum_{y_k \in Y_1} P(Y = y_k) = a \times P(Y_1)$

En fin de compte : $E(Y) \geq a \times P([Y \geq a])$

Puisque a est strictement positif, cette inégalité équivaut à l'inégalité : $\frac{E(Y)}{a} \geq P([Y \geq a])$ qui est l'inégalité de Markov : Pour tout réel a strictement positif, $\frac{E(Y)}{a} \geq P([Y \geq a])$

XX. Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire définie sur un univers Ω possédant une espérance notée E(X) et une variance notée V(X). Par définition de la variance : $V(X) = E\left[(X - E(X))^2\right]$

La variable aléatoire $(X - E(X))^2$ est une variable aléatoire positive qui, par hypothèse, a une espérance.

- On peut lui appliquer l'inégalité de Markov, en employant non pas un réel a strictement positif, mais son carré : Pour tout réel a strictement positif, $\frac{V(X)}{a^2} = \frac{E\left[(X - E(X))^2\right]}{a^2} \geq P\left(\left[(X - E(X))^2 \geq a^2\right]\right)$
- Mais puisque a est strictement positif, l'inégalité $(X - E(X))^2 \geq a^2$ est équivalente à l'inégalité : $\sqrt{(X - E(X))^2} \geq a$ autrement dit à l'inégalité $|X - E(X)| \geq a$. De ce fait : $P\left(\left[(X - E(X))^2 \geq a^2\right]\right) = P(|X - E(X)| \geq a)$

On obtient que pour tout réel a strictement positif : $\frac{V(X)}{a^2} \geq P([|X - E(X)| \geq a])$.

C'est l'inégalité de Bienaymé-Tchebychev.

Partie E : Le collectionneur.

Notons $\{a_1, a_2,..., a_n\}$ l'ensemble des animaux formant la collection. Pour q entier au moins égal à 1, considérons l'expérience consistant à répertorier les animaux obtenus au cours des q premiers achats. Cette expérience peut être modélisée par l'ensemble $\Omega_q = \{a_1, a_2,..., a_n\}^q$ des q-uplets d'éléments de $\{a_1, a_2,..., a_n\}$ muni de l'équiprobabilité (équiprobabilité car les achats sont supposés indépendants les uns des autres).

L'ensemble Ω_q ayant n^q éléments, chacune des éventualités de cet ensemble a pour probabilité $\frac{1}{n^q}$.

XXI. La variable aléatoire qui modélise le nombre d'achats nécessaires pour obtenir l'ensemble de la collection est la variable aléatoire T_n

XXII. Lorsqu'on commence une collection, le premier animal obtenu est nécessairement un nouvel animal. T_1 est la variable aléatoire constante qui prend la valeur 1 quel que soit l'élément de l'univers $\Omega_1 = \{a_1, a_2,..., a_n\}$. La variable aléatoire T_1 prend la valeur 1 avec une probabilité égale à 1.

XXIII. 1. Soit q un entier au moins égal à 2. Le collectionneur obtient toujours un même animal au cours de ses q premiers achats si et seulement si le q-uplet associé à ses achats est l'un des q-uplets $(\underbrace{a_i, a_i,..., a_i}_{q \text{ fois}})$ où $i = 1 ; 2 ;...; n$.

Il y a n q-uplets qui réalisent cet évènement, la probabilité de ce dernier est : $n \times \frac{1}{n^q} = \frac{1}{n^{q-1}}$

XXIII. 2. L'évènement « $T_2 > q$ » est réalisé si et seulement si le collectionneur obtient toujours un même animal au cours de ses q premiers achats.

Cet évènement est, pour $q \geq 2$, exactement celui dont la probabilité vient d'être calculée, mais la même formule est applicable lorsque $q = 1$ puisque $\frac{1}{n^{q-1}} = 1$ lorsque $q = 1$. Ainsi, pour tout entier q strictement positif : $P(T_2 > q) = \frac{1}{n^{q-1}}$

On note que les évènements A_q = "$T_2 > q$" sont des évènements emboîtés : pour tout entier q au moins égal à 2 : $A_q \subset A_{q-1}$

XXIII. 3. L'évènement $"T_2 = q"$ est réalisé si et seulement si l'évènement $A_{q-1} = "T_2 > q-1"$ est réalisé mais l'évènement $A_q = "T_2 > q"$ ne l'est pas. Il s'agit de l'évènement $A_{q-1} - A_q$; en raison du fait que ces évènements sont emboîtés, la probabilité de cet évènement est :

$$P(T_2 = q) = P(T_2 > q-1) - P(T_2 > q) = \frac{1}{n^{q-2}} - \frac{1}{n^{q-1}} = \frac{n-1}{n^{q-1}}$$

XXIII. 4. D'après **XXIII. 1**, s'il y a 100 animaux, avec 2 achats indépendants, la probabilité d'obtenir deux animaux identiques est exactement $\frac{1}{100}$ et celle d'obtenir deux animaux différents est exactement $\frac{99}{100}$. Il suffit d'effectuer 2 achats pour que la probabilité « d'obtenir deux animaux différents soit supérieure ou égale à 0,99 ».

XXIII. 5. La réponse est dans l'énoncé…

XXIII. 6. Si : $\begin{cases} Z_1 = T_1 \\ Z_k = T_k - T_{k-1} \quad pour \quad k \geq 2 \end{cases}$, inversement : $\begin{cases} T_1 = Z_1 \\ T_k = Z_1 + Z_2 + \ldots + Z_k \quad pour \quad k \geq 2 \end{cases}$

XXIII. 7. Sachant que le collectionneur a déjà obtenu $(k-1)$ animaux, la probabilité que ce collectionneur obtienne un nouvel animal en effectuant un achat est égale à $\frac{n-(k-1)}{n}$. La variable aléatoire Z_k modélise le nombre d'épreuves de Bernoulli identiques et indépendantes devant se succéder pour obtenir un nouvel animal. De ce fait, elle suit la loi géométrique de paramètre : $p = \frac{n-(k-1)}{n}$

On sait qu'une loi géométrique de paramètre p a pour espérance $\frac{1}{p}$ et pour variance $\frac{1-p}{p^2}$.

En l'occurrence, Z_k a pour espérance $\frac{n}{n-k+1}$ et pour variance $\frac{n \times (k-1)}{(n-k+1)^2}$.

XXIII. 8. Du fait que : $T_n = Z_1 + Z_2 + ... + Z_n = \sum_{k=1}^{n} Z_k$, et par linéarité de l'espérance :

$$E(T_n) = \sum_{k=1}^{n} E(Z_k) = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \times \sum_{k=1}^{n} \frac{1}{n-k+1}$$

Si on effectue le changement d'indice : $j = n - k + 1$, l'indice j parcourt, en décroissant, l'ensemble des entiers de 1 à n.

Ainsi : $E(T_n) = n \times \sum_{j=1}^{n} \frac{1}{j} = n \times H_n$

XXIII. 9. D'après la question **A. V**, $E(T_n) \approx n \times \ln n$.

XXIV. 1. Dans le cas de variables aléatoires mutuellement indépendantes, la variance d'une somme est égale à la somme des variances. Donc, si on admet l'indépendance mutuelle des diverses variables Z_k, la variance de T_n est égale à la somme des variances : $v(T_n) = \sum_{k=1}^{n} v(Z_k) = \sum_{k=1}^{n} \frac{n \times (k-1)}{(n-k+1)^2}$

Si on effectue le changement d'indice : $j = n - k + 1$, lorsque k parcourt l'ensemble des entiers de 1 à n, l'indice j parcourt, en décroissant, l'ensemble des entiers de 1 à n et par ailleurs : $k - 1 = n - j$.

On obtient : $v(T_n) = n \times \sum_{j=1}^{n} \frac{n-j}{j^2} = n^2 \times \sum_{j=1}^{n} \frac{1}{j^2} - n \times \sum_{j=1}^{n} \frac{1}{j}$.

Autrement dit : $v(T_n) = n^2 \times B_n - n \times H_n$

XXIV. 2. Quel que soit l'entier n strictement positif : $v(T_n) = n^2 \times B_n - n \times H_n \le n^2 \times B_n$ et compte tenu de la majoration obtenue à propos de la suite (B_n) (majoration par sa limite) dans le courant de la **partie B** : $v(T_n) = n^2 \times B_n - n \times H_n \le n^2 \times B_n \le n^2 \times \frac{\pi^2}{6}$

XXV. Soit n un entier strictement positif donné. Pour tout réel λ strictement positif, appliquons l'inégalité de Bienaymé-Tchebychev à la variable aléatoire T_n en employant le nombre strictement positif $a = \lambda\, n \ln n$:

$$\frac{V(T_n)}{(\lambda\, n \ln n)^2} = \frac{V(T_n)}{\lambda^2 \times n^2 \times (\ln n)^2} \geq P\left(\left[\left|T_n - E(T_n)\right| \geq \lambda\, n \ln n\right]\right).$$

Compte tenu de la majoration de la variance obtenue en **XXIV. 2** : $\dfrac{\pi^2}{6\,\lambda^2 \times (\ln n)^2} \geq P\left(\left[\left|T_n - E(T_n)\right| \geq \lambda\, n \ln n\right]\right)$

XXVI. Nous avons vu en question **XXIII. 8** que : $E(T_n) = n \times H_n$. Appliquons l'inégalité de Bienaymé-Tchebychev précédente en utilisant le réel $\lambda = 1$. Elle devient : $\dfrac{\pi^2}{6\,(\ln n)^2} \geq P\left(\left[\left|T_n - n \times H_n\right| \geq n \ln n\right]\right)$

Mais l'inégalité $\left|T_n - n \times H_n\right| \geq n \ln n$ est réalisée si et seulement si $T_n - n \times H_n \geq n \ln n$, c'est-à-dire $T_n \geq n \times H_n + n \ln n$ ou bien $-T_n + n \times H_n \geq n \ln n$, c'est-à-dire $n \times H_n - n \ln n \geq T_n$.

L'évènement $\left[T_n \geq n \times H_n + n \ln n\right]$ est inclus dans l'évènement $\left[\left|T_n - n \times H_n\right| \geq n \ln n\right]$ et donc a une probabilité plus faible.

On obtient : $\dfrac{\pi^2}{6\,(\ln n)^2} \geq P\left(\left[T_n \geq n \times H_n + n \ln n\right]\right)$. Pour assurer que l'inégalité $0{,}01 \geq P\left(\left[T_n \geq n \times H_n + n \ln n\right]\right)$ est réalisée, il suffit de choisir n de telle sorte que : $0{,}01 \geq \dfrac{\pi^2}{6\,(\ln n)^2}$, c'est-à-dire de telle sorte que : $(\ln n)^2 \geq \dfrac{100\,\pi^2}{6}$, soit : $n \geq \exp\left(\dfrac{10\,\pi}{\sqrt{6}}\right)$.

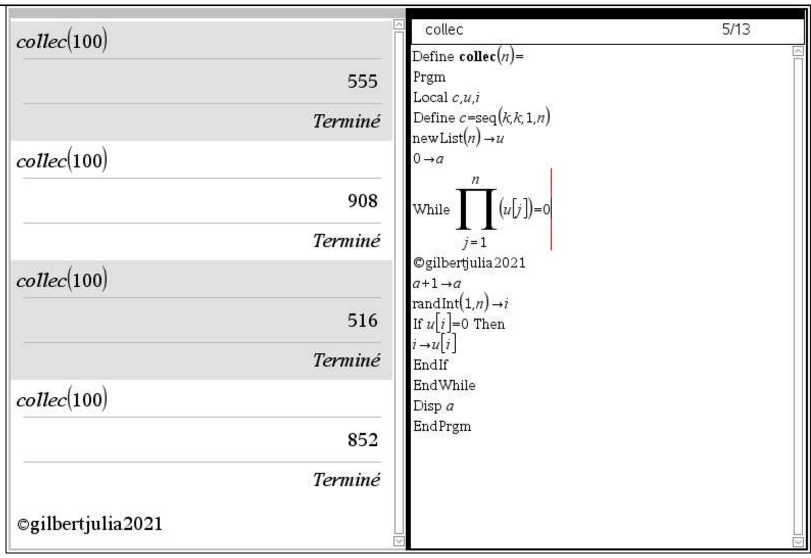
Une calculatrice indique que l'entier $n_0 = 371573$ est convenable.

Ce résultat exorbitant illustre le caractère théorique, et en pratique « anecdotique » de l'inégalité de Bienaymé-Tchebychev

Complément : Une simulation.

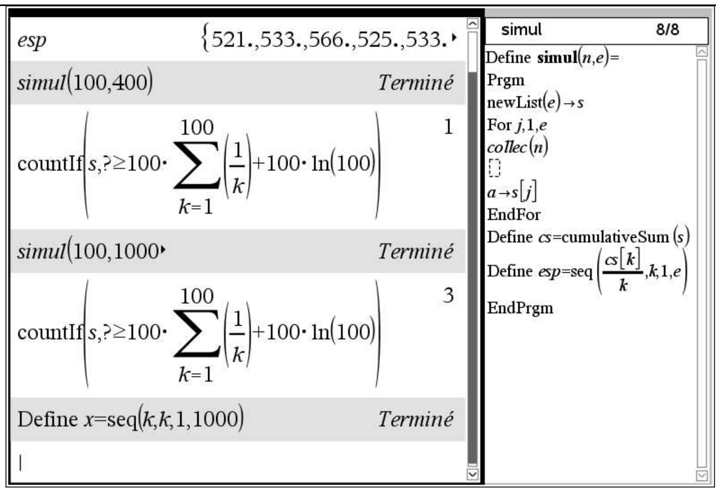
Le programme **collec** simule la confection d'une collection. Il renvoie le nombre d'achats à effectuer pour obtenir une collection complète. Il a été exécuté plusieurs fois avec 100 animaux.

Avec 100 animaux, l'espérance du nombre d'achats devrait être voisine de 519. Il semble ici que le nombre d'achats soit plus grand. Qu'en est-il vraiment ?



Le programme **simul** répète un certain nombre **e** de fois le programme précédent et répertorie dans une liste **s** les nombres d'achats nécessaires pour compléter la collection. La liste **esp** répertorie les moyennes des nombres d'achats nécessaires.

On l'a exécuté pour 400 essais puis pour 1000. On note respectivement seulement 1 cas puis 3 cas où $T_n \geq n \times H_n + n \ln n$



Il semble bien que sur le long terme, le résultat « moyenne voisine de 519 » soit plausible.