Sujet ESD2017_06. Probabilités

1. Le sujet

A. Exercice

Un restaurateur prépare chaque jour trente crèmes catalanes pour soixante-dix couverts. Le restaurateur affirme : « En moyenne deux clients sur cinq choisissent une crème catalane en dessert donc je pense que dans plus de 70% des cas j'aurais assez de crèmes catalanes »

A-t-il raison ?

B. Les réponses de deux élèves de première scientifique

Élève 1

J'ai reconnu la loi binomiale : $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ *avec* $n=70,\ p=0{,}4$ *et* $k=30$.

La probabilité est donc environ 0,085.
Donc le restaurateur a tort, il satisfait seulement 8,5% de la demande. Cela me semble peu.

Elève 2

J'ai écrit un algorithme qui calcule le nombre de crèmes catalanes commandées par les soixante-dix clients puis je l'ai répété mille fois pour avoir une moyenne :

```
pour I variant de 1 à 1000 faire
|   pour J variant de 1 à 70 faire
|   |   Affecter à aléa une valeur choisie au hasard parmi 1, 2, 3, 4 ou 5.
|   |   si aléa < 3 alors
|   |   |   C prend la valeur C + 1
|   |   fin
|   fin
fin
M prend la valeur C/1000
Afficher M
```

J'ai lancé 3 fois l'algorithme et j'ai trouvé 27,9 ; 28,1 et 28. En moyenne 28 crèmes catalanes sont commandées par les 70 clients donc ils seront tous satisfaits.

C. Le travail à exposer devant le jury

1. Analysez les productions de chaque élève en mettant en évidence leurs réussites et leurs éventuelles erreurs, ainsi que l'accompagnement que vous pourriez leur proposer pour les aider à progresser.

2. En vous appuyant sur les productions des élèves, présentez une correction de l'exercice telle que vous l'exposeriez devant une classe de première scientifique.

3. En motivant vos choix, proposez deux exercices sur le thème *probabilités* à des niveaux de classe différents.

2. Eléments de correction

Cet exercice porte sur la notion de loi binomiale, notamment sur sa distribution. Contrairement à son apparence, il ne s'agit pas du tout de « l'utilisation de la loi binomiale pour une prise de décision à partir d'une fréquence » (il n'y a dans l'énoncé aucune fréquence observée). La loi binomiale n'est pas un outil pour l'exercice mais l'objet même de l'exercice.

Le nombre X de crèmes catalanes commandées par les clients est par hypothèse une variable aléatoire qui suit la loi $_{gjulia2017}$ $B(70\ ;\ 0{,}4)$, d'espérance 28 et d'écart type 4,1. Le restaurateur est en rupture de stock si et seulement si $X \geq 31$, (il est en rupture de stock lorsqu'il a servi déjà 30 crèmes catalanes et qu'au moins un client lui en demande une de plus).

On s'intéresse donc dans cet exercice aux probabilités $P([X \leq 30])$ et $P([X \geq 31]) = 1 - P([X \leq 30])$ qu'il faudra calculer d'une façon ou d'une autre. L'affirmation du restaurateur est là pour lancer l'exercice en une sorte de défi à relever (valider ou invalider cette affirmation).

Les paramètres 28 et 4,1 ne nous aident pas vraiment. Certes, on peut estimer « *a visto de nas* » que, la demande se situant entre 24 et 32 lors d'environ 66 % des journées, l'affirmation du restaurateur semble « plausible ». Mais sa véracité reste entièrement à démontrer.

1. Analyse de travaux d'élèves.

Elève 1.
Solution incorrecte.
Cet élève reconnaît explicitement la loi binomiale à laquelle il faudra faire référence. Il a su traiter l'information donnée par l'énoncé, la traduire mathématiquement et identifier la notion en jeu. Pour preuve, il calcule de façon correcte la probabilité $_{gj}$ $P([X = 30])$.

Cet élève a cependant une compréhension incorrecte de la situation, notamment à propos de l'interprétation qu'il se donne de «… j'aurais assez de … ». Il confond l'évènement « le restaurateur sert exactement 30 crèmes catalanes » et l'évènement « le restaurateur sert au plus 30 crèmes catalanes ».

Il émet un doute sur la pertinence de sa réponse, mais sans que ce doute l'amène à reconsidérer sa résolution.

Il faut bien entendu conforter ce doute et amener cet élève à analyser l'énoncé : « Pourquoi cela te semble peu (que penser d'un restaurateur qui satisfait moins d'un client sur dix …) ? Quel est l'évènement dont tu as calculé la probabilité (amener cet élève a verbaliser ce qu'il a calculé) ? Est-ce bien la même chose que l'évènement « le restaurateur a prévu suffisamment de crèmes catalanes » ?

Chouquerouste.
Solution incorrecte.
Chouquerouste a une bonne compréhension de la situation à étudier. Il a pour intention clairement exprimée de construire une simulation d'une série de 1000 journées.

Cependant, pour définir l'objectif de sa simulation, il se base sur un théorème-élève incorrect. Selon lui : « $x - E(X)$ *assez grand* $\Rightarrow P([X > x]) = 0$ », le nombre $E(X)$ étant estimé empiriquement par une moyenne à l'aide d'une simulation (on ne peut pas lui tenir rigueur de ne pas fournir de « preuve » que la moyenne théorique peut être convenablement approchée par une moyenne observée : on étudie au lycée la fluctuation d'une fréquence, non celle d'une moyenne).

À cet effet, il rédige un algorithme comportant deux boucles imbriquées et un test conditionnel de sélection. Il semble maîtriser assez bien une telle rédaction.

Cependant, le compteur **c** ne dénombre pas parmi ces 1000 journées celles où le restaurateur est en rupture de stock (ce qui était son intention) mais compte[1] le nombre de crèmes catalanes commandées par les 70000 clients.

$\dfrac{c}{1000}$ représente alors le nombre de crèmes catalanes commandées en moyenne par 70 des 70000 clients. C'est ce qui explique que les trois moyennes obtenues sont toutes très proches de l'espérance de X.

Il aurait fallu que le compteur soit remis à zéro avant chaque « journée » et qu'un test conditionnel permette de distinguer le cas où le nombre de crèmes commandées est au plus 30 de celui où il est au moins 31.

Il paraît difficile d'aider Chouquerouste à « améliorer » (?) sa démarche, sauf à s'engager dans des complications disproportionnées. Il semble plus judicieux de l'aider à prendre conscience de la nature de la loi en jeu, à savoir $B(70\ ;\ 0{,}4)$, dont la moyenne est bien 28 mais dont la dispersion n'a pas été mesurée et mériterait de l'être. On peut attirer son attention sur le fait que, selon sa vision de la situation, la prévision « 70 % » aurait aussi bien pu être « 100 % » et cela indépendamment de la dispersion des résultats. Etrange, non ?

2. Une correction de l'exercice.

C'est bien la production de l'élève 1 qui peut servir du mieux possible de point de départ à une correction.

On fait apparaître que le nombre X de crèmes catalanes commandées par les clients est par hypothèse une variable aléatoire qui suit la loi $B(70\ ;\ 0{,}4)$ dont on fait préciser moyenne et écart type (ce qui permet d'avoir une idée de la répartition des valeurs). Le restaurateur est en rupture de stock si et seulement si $X \geq 31$, (il est en rupture de stock lorsqu'il a servi déjà 30 crèmes catalanes et qu'au moins un client lui en demande une de plus). Il s'agit de savoir s'il est vrai ou faux que $P([X \leq 30]) \geq 0{,}7$. Le restaurateur affirme que oui, mais il nous faut le vérifier.

L'élève 1 a calculé seulement $P([X = 30])$, mais $P([X \leq 30]) = P([X = 0]) + P([X = 1]) + \ldots + P([X = 30])$. Il faut calculer la somme de toutes ces probabilités. L'usage d'un tableur ou d'une calculatrice permettra d'automatiser ce calcul.

La copie d'écran ci-contre affiche en surbrillance à la fois le résultat $P([X = 30])$ obtenu par l'élève 1 et la somme des probabilités $\sum_{k=0}^{30} P([X = k])$, c'est à dire $P([X \leq 30])$.

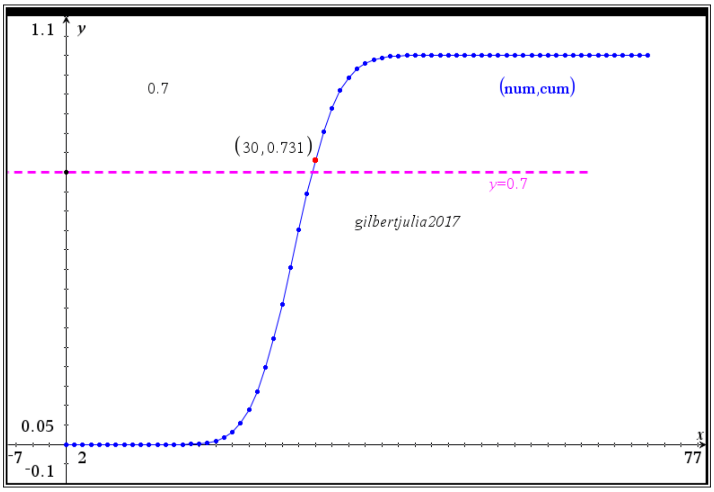
Est affichée également en cellule **d31** la probabilité de rupture de stock.

	A num	B prb	C cum	D	E
=	=seq(k,k	=seq(ncr(70,k)*(0.4)^k*(0.6)^(70−	=cumulativesum(prb)		
28	27.	0.094711	0.454711		
29	28.	0.096966	0.551677		
30	29.	0.093622	0.645299	gilbertjulia2017	
31	30.	0.0853	0.730599	0.269401	
32	31.	0.073376	0.803976		
33	32.	0.059618	0.863594		
34	33.	0.045768	0.909362		
35	34.	0.033204	0.942566		
36	35.	0.022768	0.965334		
37	36.	0.014757	0.980091		
38	37.	0.009041	0.989132		
39	38.	0.005234	0.994366		
40	39.	0.002863	0.997229		

[1] Sacré Chouquerouste ! Le voici pris en flagrant délit de récidive. L'an dernier, il avait commis dans la même classe la même erreur (voir ESD 2016_12) à peu près dans les mêmes circonstances. Comme quoi, preuve que le redoublement ne sert à rien puisque les mêmes élèves commettent les mêmes erreurs aux mêmes endroits.

Eventuellement, une représentation graphique du nuage de points $\left(k\ ;\ \sum_{j=0}^{j=k} P([X=j])\right)$ met elle aussi en évidence que $P([X \leq 30]) \geq 0{,}7$.

On peut donner raison au restaurateur.



En revanche, il n’y a à mon avis pas grand-chose à tirer de la production de Chouquerouste, sinon à se fourvoyer assez loin du sujet.

Si on décide quand même d’exploiter sa production, on proposera un algorithme rectifié (insister sur l’importance du positionnement de la saisie de la valeur initiale d’une variable dans un algorithme comportant une ou plusieurs boucles). Et il est raisonnable de s’en tenir là …

Sinon, pour information, il y aurait deux exploitations possibles :

- Ou bien on pourrait s'en servir pour valider le résultat théorique $P([X>30])=0{,}2694$ que l'on a déjà obtenu comme indiqué ci-dessus. Dans ce cas, on vérifie si une fréquence observée est bien dans un intervalle de fluctuation du type $\left[0{,}2694-\frac{1}{\sqrt{n}}\ ;\ 0{,}2694+\frac{1}{\sqrt{n}}\right]$. On reste ainsi strictement dans le cadre du programme de première.
- Ou bien on pourrait s'en servir pour tenter directement de prouver que le restaurateur a raison. La probabilité de rupture de stock reste inconnue, on ne la calcule pas. On cherche à l'aide d'une simulation à situer cette probabilité inconnue $P([X>30])$ dans un intervalle de confiance ne contenant pas 0,3. Mais dans ce cas, ce sont des figues d'un autre panier, on sort carrément du programme de première.

Exercices complémentaires

Voir chapitre 8, pages 180 à 203 de REDCM, mais pas seulement puisque les programmes ont changé depuis, notamment au niveau du collège.

3. Commentaire

Le dessert connu sous le nom de « crème catalane » est à l'origine la « crema de Sant Josep ». Sa recette originale daterait du XIVième siècle. Son succès actuel est dû au fait que cette crème catalane est devenue un best-seller de la cuisine industrielle internationale.

Il serait intéressant de savoir si le restaurateur en question dans l'énoncé fait sa crème catalane lui-même ou bien l'achète surgelée. Au moins autant, pour peu que l'on soit un tantinet gourmet, que de savoir si oui ou non il restera de ce dessert en arrivant le dernier dans ce restaurant.

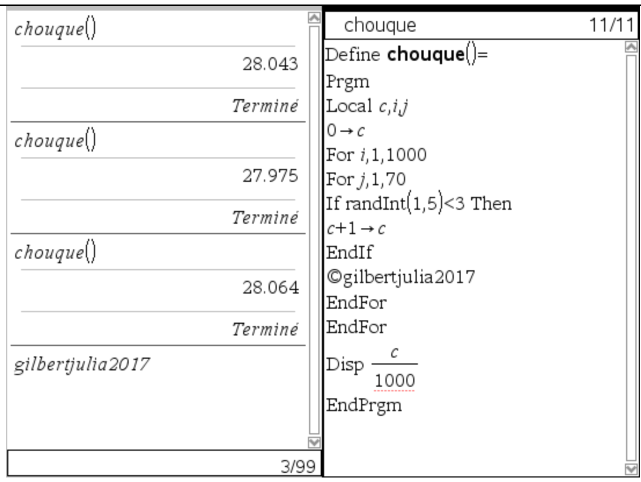
Si c'est pour une crème surgelée, inutile de se presser. En effet, la crème catalane industrielle et la « crema de Sant Josep » faite localement par quelques estaminets sud-catalans traditionnels sont deux desserts aussi dissemblables qu'une carpe et un lapin. Il faut avoir goûté une « crema de Sant Josep », saupoudrée devant soi de cassonade, marquée au fer rouge tiré du fourneau et qui fait grésiller le caramel jusqu'au brun clair, pour s'en rendre compte. Pour un tel dessert, alors oui cela vaut le coup d'être dans les 30.

4. Pour aller plus loin

Pour information, voici deux exemples d'algorithmes Chouquerouste, original puis modifié, d'abord avec *n*SpireCAS puis avec Algobox, logiciel lui aussi condamné.

Ci-contre, le programme **chouque()** reprend exactement l'algorithme proposé par Chouquerouste :

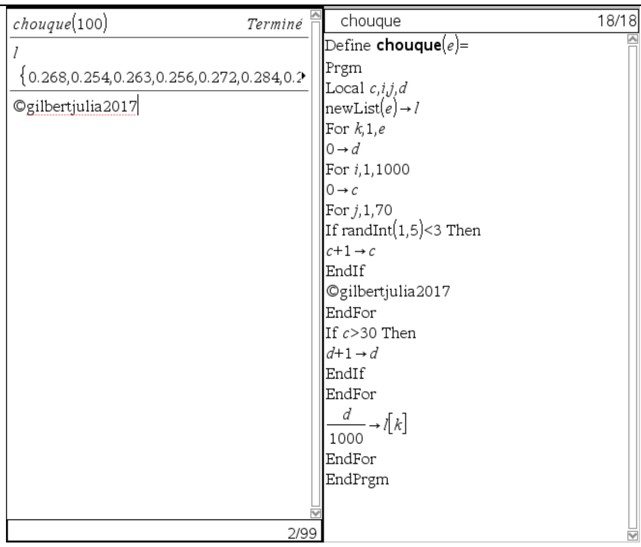
On observe une faible dispersion des résultats $\frac{c}{1000}$ autour de 28, moyenne du nombre de demandes de crème catalane sur 70 desserts commandés parmi 70000.



Le voici maintenant modifié de façon à sélectionner les journées de ruptures de stock.

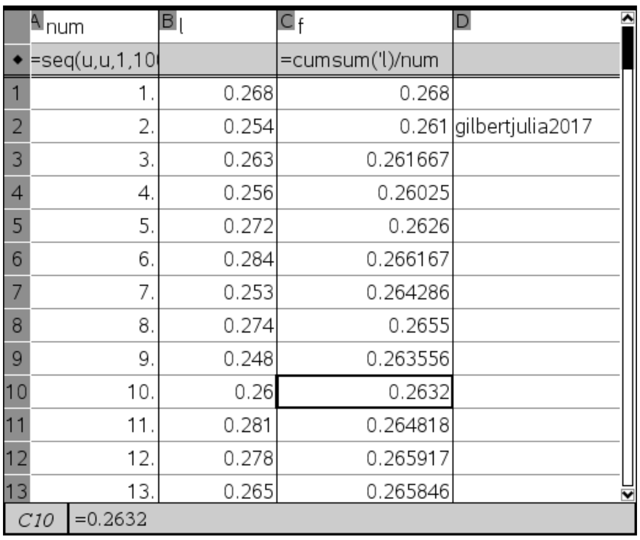
La variable **c** n'est pas initialisée au même endroit et une autre variable comptabilise ces journées où il y a rupture de stock.

On dresse une liste de résultats pour une éventuelle exploitation ultérieure (que j'ai finalement abandonnée …). Chacun des termes $l[k]$ de la liste **l** correspond à un résultat que Chouquerouste avait, semble-t-il, l'intention d'obtenir.



On peut conjecturer en examinant la liste **l** que la fréquence des journées de rupture de stock est inférieure à 0,3 et même est inférieure à 0,28.
La fréquence $f[10]=0{,}2632$ par exemple est une fréquence de rupture de stock parmi 10000 journées.

- Si on a déjà calculé $P([X>30])$, on vérifie que $f[10]=0{,}2632$ est bien dans l'intervalle de fluctuation $[0{,}2594\ ;\ 0{,}2794]$.
- Si $P([X>30])$ est supposé inconnu, on peut affirmer, avec une probabilité au moins égale à 95 %, que $P([X>30])$ est dans l'intervalle de confiance $[0{,}2532\ ;\ 0{,}2732]$.



	A num	B l	C f	D
◆	=seq(u,u,1,10		=cumsum('l)/num	
1	1.	0.268	0.268	
2	2.	0.254	0.261	gilbertjulia2017
3	3.	0.263	0.261667	
4	4.	0.256	0.26025	
5	5.	0.272	0.2626	
6	6.	0.284	0.266167	
7	7.	0.253	0.264286	
8	8.	0.274	0.2655	
9	9.	0.248	0.263556	
10	10.	0.26	0.2632	
11	11.	0.281	0.264818	
12	12.	0.278	0.265917	
13	13.	0.265	0.265846	

C10 =0.2632

Une étude avec Algobox amènerait à des conclusions identiques.

Algorithme Chouquerouste

```
VARIABLES
  i EST_DU_TYPE NOMBRE
  j EST_DU_TYPE NOMBRE
  a EST_DU_TYPE NOMBRE
  c EST_DU_TYPE NOMBRE
  m EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  c PREND_LA_VALEUR 0
  //gilbertjulia2017
  POUR i ALLANT_DE 1 A 1000
    DEBUT_POUR
    POUR j ALLANT_DE 1 A 70
      DEBUT_POUR
      a PREND_LA_VALEUR random()
      SI (a<0.4) ALORS
        DEBUT_SI
        c PREND_LA_VALEUR c+1
        FIN_SI
      FIN_POUR
    FIN_POUR
  m PREND_LA_VALEUR c/1000
  AFFICHER m
FIN_ALGORITHME
```

Faible dispersion des résultats autour de 28

Algorithme modifié

```
VARIABLES
  i EST_DU_TYPE NOMBRE
  j EST_DU_TYPE NOMBRE
  c EST_DU_TYPE NOMBRE
  m EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  //gilbertjulia2017
  m PREND_LA_VALEUR 0
  POUR i ALLANT_DE 1 A 1000
    DEBUT_POUR
    c PREND_LA_VALEUR 0
    POUR j ALLANT_DE 1 A 70
      DEBUT_POUR
      SI (random()<0.4) ALORS
        DEBUT_SI
        c PREND_LA_VALEUR c+1
        FIN_SI
      FIN_POUR
    SI (c>30) ALORS
      DEBUT_SI
      m PREND_LA_VALEUR m+1
      FIN_SI
    FIN_POUR
  m PREND_LA_VALEUR m/1000
  AFFICHER m
FIN_ALGORITHME
```

Plus grande dispersion des résultats autour de 0,2694