ESD 2015 – 14 : Prise de décision

1. Le sujet

A. L'exercice proposé au candidat

Pour chacune des propositions suivantes, dire si la proposition est vraie ou fausse en justifiant la réponse.

L'entreprise MICRO vend en ligne du matériel informatique, notamment des ordinateurs portables et des clés USB.

Partie A

L'autonomie de la batterie qui équipe les ordinateurs portables distribués par la société MICRO, exprimée en heures, suit une loi normale d'espérance $\mu = 2$ et d'écart-type $\sigma = 2$

Proposition 1 :

La probabilité que l'ordinateur ait une autonomie supérieure ou égale à 10 h est inférieure à 0,2.

Partie B

L'entreprise MICRO vend également des clés USB et communique sur ce produit en affirmant que 98% des clés commercialisées fonctionnent correctement. Sur 1000 clés prélevées dans le stock, 50 clés se révèlent défectueuses.

Proposition 2 :

Ce test, réalisé sur ces 1000 clés, ne remet pas en cause la communication de l'entreprise.

D'après un sujet de baccalauréat

B. Les réponses de trois élèves de terminale S à la partie B

Élève 1

Il y a 950 clés qui fonctionnent correctement, ce qui fait 95% : c'est moins que ce que dit l'entreprise, mais on peut dire qu'on n'a pas eu de chance avec le lot car ça n'est pas très loin de 98%.

Élève 2

J'ai préparé sur tableur une simulation en entrant =SI(ALEA()<0,98;1;0) , que j'ai recopié pour avoir 1000 clés et j'ai compté le nombre de fois où on obtient 1 : en relançant le calcul, cela varie, mais tout le temps au dessus de 950 : on peut donc penser qu'il y a un problème dans la communication.

Élève 3

J'utilise l'intervalle vu en seconde $I = \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right] = \left[0,98 - \frac{1}{\sqrt{1000}} \; ; \; 0,98 + \frac{1}{\sqrt{1000}}\right] = [0,948 \; ; \; 1,012]$

Il n' y a donc pas de problème.

C. Le travail à exposer devant le jury

1. Analysez la capacité de chaque élève à s'engager dans une démarche de recherche.

2. Exposez une correction de l'exercice comme vous le feriez avec une classe de terminale scientifique en appuyant éventuellement votre propos sur une simulation.

3. Proposez deux exercices sur le thème *prise de décision* à des niveaux de classe différents. Vous motiverez vos choix en indiquant les compétences que vous cherchez à développer chez les élèves.

2. Eléments de correction

L'exercice porte sur la notion de loi normale et ses applications en échantillonnage. Il est paraît-il tiré d'annales de baccalauréat, et cela se ressent : les deux « parties » de l'exercice sont indépendantes et n'ont de rapport l'une avec l'autre que l'habillage, quelque peu cousu de fil blanc.

La partie A porte sur une propriété de la distribution d'une loi normale : « connaître une valeur approchée de la probabilité de l'évènement $X \in [\mu - \sigma \ ; \ \mu + \sigma]$ » (savoir)

La partie B porte sur la notion de prise de décision. Il s'agit de « connaître l'intervalle de fluctuation asymptotique au seuil de 95 % » (savoir) et l'utiliser en tant qu'aide à la décision (savoir faire).

La question 2 posée par le jury peut légitimement déconcerter, car l'exercice proposé n'est pas un exercice de recherche, mais plutôt un exercice d'évaluation de connaissances. Il n'y a pas à proprement parler de « démarche de recherche » à enclencher.

1. *Elève 1.*
Cet élève s'en tient à une idée intuitive. Il a une conception rudimentaire de la notion de pourcentage, car il compare 95 et 98 non pas comme étant des pourcentages, mais comme étant des nombres entiers naturels. Il considère que ces deux entiers naturels ne sont « pas très loin » (comme il dirait qu'une caisse contenant 95 pommes et une autre contenant 98 pommes ne sont « pas très loin » d'être autant remplies). De ce fait, il n'a aucune raison de s'engager dans une activité de recherche, quelle qu'elle soit.
Sa démarche est incorrecte.
Il faudrait faire calculer à cet élève les pourcentages de clés défectueuses. Est-ce que 5 % de clés défectueuses c'est « pas très loin » de 2 % ?

Elève 2.
Cet élève s'engage dans une démarche dont le principe a été vu en seconde. Il a l'intention de comparer la proportion de clés défectueuses observées à celles d'un échantillon issu d'une simulation d'un certain nombre de tests sur 1000 prélèvements (on ne sait pas combien, mais comme il a « relancé le calcul », ce nombre de tests doit être de l'ordre de la dizaine).
Cet élève a su « construire une simulation numérique prenant appui sur la modélisation » mais sans pour autant aller au bout de la démarche car il ne fait référence à aucun critère permettant de justifier l'acceptation ou le rejet de la décision de remise en cause.
Il apparaîtra utile de fixer des normes pour uniformiser la validité de la conclusion. Il faut cependant reconnaître que 950 est très excentré par rapport aux valeurs usuelles attendues. Cet élève n'a pas tout à fait tort d'affirmer qu'on obtient « tout le temps » plus de 950.
On peut considérer sa démarche comme « inaboutie » en raison de l'absence de critères fiables pour mesurer l'excentricité de la valeur 950 par rapport aux valeurs issues de la simulation considérées comme normales.

Elève 3
Cet élève identifie correctement le type de situation à traiter : prendre une décision à partir d'un échantillon et pour cela construire un intervalle de fluctuation au seuil de 95 %. Il « recherche » … dans sa mémoire !

Il se souvient qu'un tel intervalle de fluctuation a été défini en seconde. Mais cet élève ne prend pas garde qu'il est en dehors du domaine de validité dans lequel cet intervalle de fluctuation est opérationnel. Si la taille de l'échantillon est supérieure à 25, en revanche la probabilité p n'est pas dans l'intervalle $[0{,}2 \ ; \ 0{,}8]$. La méthode de seconde est inapplicable.
Sa démarche est de ce fait incorrecte.

La production de cet élève est cependant intéressante, car elle va permettre de rafraîchir quelques mémoires et de souligner les avantages de l'intervalle de fluctuation asymptotique tel qu'il est présenté en terminale.

En résumé, seul l'élève 2 a fait preuve en un certain sens d'une « démarche de recherche », mais un tel type de démarche n'était pas du tout l'objectif visé par l'exercice. Les deux autres élèves n'avaient aucune raison de « rechercher ».

2. ***Correction de l'exercice.***

Partie A.

Les élèves doivent retenir que de façon générale, étant donnée une variable aléatoire X de moyenne m et d'écart type s, la probabilité que X soit dans l'intervalle $[m-s\ ;\ m+s]$ est environ $\dfrac{2}{3}$

Si plus particulièrement X suit une loi normale $N(\mu,\ \sigma)$, loi dont la répartition de part et d'autre de la moyenne est symétrique, alors par-dessus le marché on dispose d'une table :

À 10^{-4} près : $P(\mu-\sigma \le X \le \mu+\sigma)=0{,}6826$ et $P(X \le \mu-\sigma)=P(X \ge \mu+\sigma)=0{,}1587$

Ainsi : $P(\mu-\sigma \le X \le \mu+\sigma) \ge \dfrac{2}{3}$ et $_{gjulia2015}$ $P(X \le \mu-\sigma)=P(X \ge \mu+\sigma) \le \dfrac{1}{6}$

La proposition 1 est exacte.

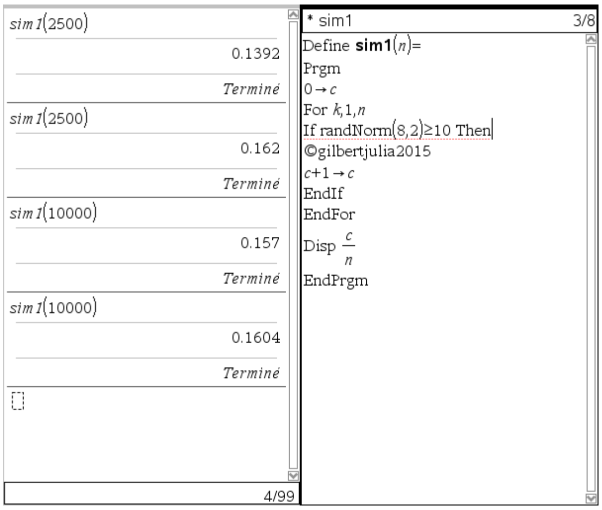
Simulation possible :

En supposant inconnue la probabilité $P(X \ge \mu+\sigma)$, une simulation peut l'estimer par un intervalle de confiance dans le cas qui nous intéresse :

Le petit programme ci-contre simule une suite de n valeurs d'une variable aléatoire X suivant la loi normale $N(\mu,\ \sigma)$ et un compteur c dénombre les valeurs de X qui sont supérieures ou égales à 10. On l' exécuté quatre fois, deux fois avec $n=2500$ et deux fois avec $n=10000$. Chacune des exécutions détermine un intervalle de confiance au seuil de 95 % dans lequel se trouve $P(X \ge 10)$.

On obtient successivement $[0{,}1192\ ;\ 0{,}1592]$ puis $[0{,}142\ ;\ 0{,}182]$; $[0{,}147\ ;\ 0{,}167]$; $[0{,}1504\ ;\ 0{,}1704]$.

Chaque intervalle contient la valeur 0,1587 donnée par la table (l'un de très peu il est vrai) et tous excluent les valeurs égales ou supérieures à 0,2.



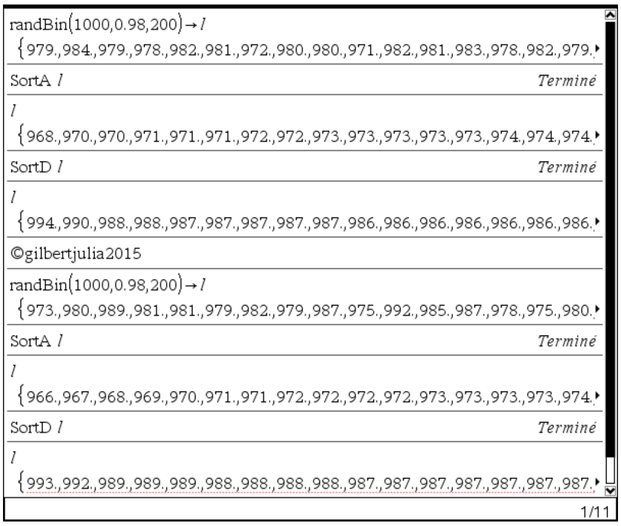
Partie B.

On s'appuie sur les productions des élèves 2 et 3.

L'élève 2 dit que l'on obtient « tout le temps » plus de 950 et effectivement, il n'a pas tout à fait tort.

Exemples de simulation « Elève 2 »

Une simulation portant sur 200 tests de 1000 prélèvements montre que les fréquences obtenues devraient être concentrées à 95 % dans l'intervalle $[0{,}973\ ;\ 0{,}986]$ ou $[0{,}972\ ;\ 0{,}987]$ (résultats fournis respectivement par l'une et l'autre des deux simulations ci-contre).



L'élève 3 donne quant à lui $[0{,}948\ ;\ 1{,}012]$ comme intervalle de fluctuation de la fréquence au seuil de 95 %. Si c'est exact, on devrait trouver quand même de temps à autre (dans environ 2,5 % des cas) des valeurs plus petites que 948. Or, c'est loin d'être le cas dans la simulation « Elève 2 ». Il y a donc un problème …

- On fait remarquer que la méthode de seconde est inutilisable car p n'appartient pas à l'intervalle $[0{,}2\ ;\ 0{,}8]$, condition pour considérer la méthode comme opérationnelle. C'est peut-être ce qui explique le paradoxe précédent. De toute façon, il faut trouver autre chose.
- On fait remarquer aussi que la méthode « Elève 3 » n'est pas pleinement satisfaisante car, chacun effectuant sa simulation, l'intervalle de fluctuation obtenu n'est pas le même pour tout le monde.

On rappelle alors l'intervalle de fluctuation asymptotique au seuil 95 % défini en terminale :

$$J=\left[p-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\,p+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]_{gjulia2015}=\left[0{,}98-\frac{0.2744}{\sqrt{1000}}\ ;\ 0{,}98+\frac{0{,}2744}{\sqrt{1000}}\right]=[0{,}971\ ;\ 0{,}989]$$

Son domaine de validité est usuellement $n\geq 30\ ;\ n\,p\geq 5\ ;\ n(1-p)\geq 5$, critères qui sont tous satisfaits ici.

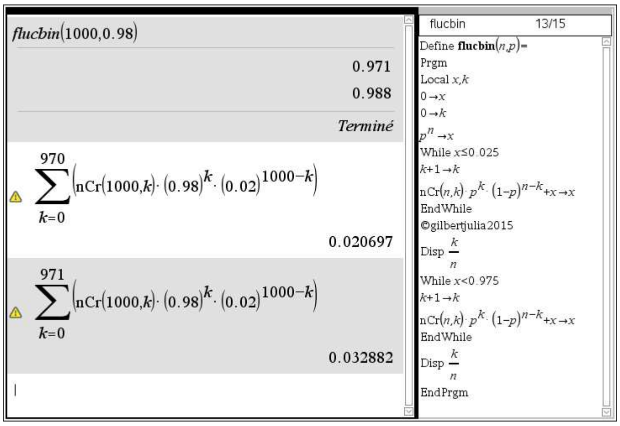
La valeur 0,95 est nettement en dehors de cet intervalle. Le test remet en cause la communication de l'entreprise.

On met l'accent sur la meilleure performance par rapport à la méthode de seconde, en particulier pour des valeurs de p faibles ou fortes (comme 0,98 où la méthode seconde est invalide). On emploie le coefficient 0,2744 là où en seconde on employait 1. On note aussi une compatibilité beaucoup plus forte avec les intervalles de fluctuation déduits de simulations. Son avantage est que cet intervalle est indépendant de l'opérateur qui met en œuvre le test.

3. Commentaires

1. Pour mémoire, en première on aurait construit un intervalle de fluctuation à l'aide de la loi binomiale comme ci-contre.

Cet intervalle est très proche de l'intervalle de fluctuation asymptotique de la classe de terminale.



2. La partie B tourne autour de la notion d'intervalle de fluctuation asymptotique, intervalle de la forme $\left[p-u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\,p+u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ où le coefficient u_α dépend du seuil que l'on se fixe. Ou bien on connaît ce type d'intervalle ou bien on ne le connaît pas, on ne peut pas l'inventer. Il n'y a pas de « démarche de recherche » à mettre en oeuvre pour prendre une décision, mais un protocole à suivre.