Quelques remarques et propositions sur les sujets de bac S 2013

Probabilités, statistiques et algorithmique

Hubert Raymondaud(*)

L'article qui suit a été fait à partir de la synthèse de remarques personnelles et de celles de collègues de l'APMEP et de la Commission Inter IREM Statistique et Probabilités.

Ces remarques portent sur quelques erreurs importantes et sur des considérations plus générales sur l'esprit et la rédaction des exercices concernant les probabilités, la statistique et l'algorithmique.

J'aborderai le problème posé par le calcul de probabilités binomiales effectué par l'approximation par la loi normale. Je commenterai la confusion entre la variable aléatoire F, fréquence de succès dans un schéma de Bernoulli, et la valeur numérique f obtenue lors d'une réalisation de ce schéma. Je commenterai l'utilisation des tables et du vocabulaire des valeurs seuil proposée dans les sujets. Je terminerai par quelques remarques sur les exercices d'algorithmique.

J'illustrerai certaines de mes propositions par des algorithmes mis en œuvre avec des lignes de commandes ou des programmes en langage **R**. Ce langage est utilisé dans toutes les universités enseignant les probabilités et la statistique et dans les organismes de recherche (INSERM, INRA, CNRS, MNHN, CIRAD, …). Le lecteur désireux de faire connaissance avec **R** et d'avoir quelques exemples de ce que l'on peut en faire en algorithmique et en mathématique, au lycée peut consulter « Quelques activités avec **R**, …, illustrations en analyse, probabilité et statistique(s) au lycée. » (http://revue.sesamath.net/spip.php?article501).

A – Utilisation de l'approximation gaussienne de la loi binomiale pour calculer des probabilités binomiales

Le programme ne mentionne pas cette pratique. Les précautions qui sont préconisées dans les ouvrages classiques (correction de continuité, $\{n \geq 30$ et $np \geq 5$ et $n(1-p) \geq 5\}$ …) sont parfois insuffisantes, comme je le montre à la fin de cette partie. Elles sont devenues complètement inutiles dans le cas de l'utilisation de tableurs, de logiciels ou de librairies spécialisés, qui possèdent des procédures d'approximation numérique très performantes.

Dans tous les sujets du baccalauréat S 2013, cette approximation était inutile puisque les calculs pouvaient être faits avec les fonctions spécialisées des calculatrices d'entrée de gamme de lycée (sauf peut-être celles des marques de distributeurs, toutefois très peu répandues).

(*) LEGTA Louis Giraud à Carpentras-Serres. hubert.raymondaud@educagri.fr;

Le document d'accompagnement se positionne clairement contre :

Remarque

Quand $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, il est courant de faire les calculs impliquant une variable binomiale en la remplaçant par une variable suivant une loi normale de mêmes espérance et variance.

Seul le programme de STI2D-STL mentionne cette pratique, qui ne doit donc pas être mise en œuvre dans les autres filières où tous les calculs de probabilités se font à la calculatrice en utilisant la loi exacte (au programme), quelle qu'elle soit.

Les calculs d'intervalles de fluctuation et d'intervalles de confiance se font avec les formules données dans le programme.

Ministère de l'éducation nationale, de la jeunesse et de la vie associative (DGESCO) **Page 21 sur 70**
Mathématiques - Probabilités et statistique
http://eduscol.education.fr/prog

Plusieurs sujets demandaient de faire cette approximation :

► **Pondichéry, exercice 4 question 3. b) :**

3. Cette entreprise emploie 220 salariés. Pour la suite on admet que la probabilité pour qu'un salarié soit malade une semaine donnée durant cette période d'épidémie est égale à $p = 0{,}05$.

On suppose que l'état de santé d'un salarié ne dépend pas de l'état de santé de ses collègues.

On désigne par X la variable aléatoire qui donne le nombre de salariés malades une semaine donnée.

a) Justifier que la variable aléatoire X suit une loi binomiale dont on donnera les paramètres.
Calculer l'espérance mathématique μ et l'écart type σ de la variable aléatoire X.

b) On admet que l'on peut approcher la loi de la variable aléatoire $\frac{X-\mu}{\sigma}$ par la loi normale centrée réduite c'est-à-dire de paramètres 0 et 1.
On note Z une variable aléatoire suivant la loi normale centrée réduite. Le tableau suivant donne les probabilités de l'événement $Z < x$ pour quelques valeurs du nombre réel x.

x	−1,55	−1,24	−0,93	−0,62	−0,31	0,00	0,31	0,62	0,93	1,24	1,55
$P(Z < x)$	0,061	0,108	0,177	0,268	0,379	0,500	0,621	0,732	0,823	0,892	0,939

Calculer, au moyen de l'approximation proposée en question b), une valeur approchée à 10^{-2} près de la probabilité de l'évènement : « le nombre de salariés absents dans l'entreprise au cours d'une semaine donnée est supérieur ou égal à 7 et inférieur ou égal à 15 ».

• L'énoncé propose de faire l'approximation et propose de passer par une variable centrée réduite. De plus l'approximation est inutile puisque l'on peut calculer cette probabilité directement à la calculatrice qui donne $P(7 \leq X \leq 15) \approx 0{,}839$.
L'approximation par la loi normale donne $P(7 \leq X \leq 15) \approx 0{,}784$. L'approximation est mauvaise, faute de correction de continuité.
Il est donc impossible de respecter la précision demandée, à 10^{-2} près.

► **Antilles Guyanne, exercice 2 partie B question 2. c) :**

2. Pour estimer r, on interroge 400 personnes et on note X la variable aléatoire comptant le nombre de bonnes réponses.

On admettra qu'interroger au hasard 400 étudiants revient à effectuer un tirage avec remise de 400 étudiants dans l'ensemble de tous les étudiants.

a. Donner la loi de X et ses paramètres n et p en fonction de r.

b. Dans un premier sondage, on constate que 240 étudiants répondent A, parmi les 400 interrogés.
Donner un intervalle de confiance au seuil de 95 % de l'estimation de p.
En déduire un intervalle de confiance au seuil de 95 % de r.

c. Dans la suite, on suppose que $r = 0,4$. Compte-tenu du grand nombre d'étudiants, on considérera que X suit une loi normale.

i. Donner les paramètres de cette loi normale.

ii. Donner une valeur approchée de $\mathrm{P}(X \leqslant 250)$ à 10^{-2} près.
On pourra s'aider de la table donnée ci-dessous, qui donne une valeur approchée de $\mathrm{P}(X \leqslant t)$ où X est la variable aléatoire de la question 2. c.

•L'énoncé propose l'approximation d'une loi binomiale par une loi normale alors que, dans le programme, le théorème de Moivre-Laplace passe par une variable centrée réduite. Là encore, l'approximation est inutile puisque l'on peut calculer cette probabilité directement à la calculatrice qui donne $\mathrm{P}(X \leq 250) \approx 0{,}858$.
L'approximation par la loi de Gauss donne $\mathrm{P}(X \leq 250) \approx 0{,}846$. Là encore, on n'atteint pas la précision demandée, à 10^{-2} près.
Nous avons donc demandé à l'IGEN de préciser si cette approximation est exigible et dans ce cas de préciser les conditions dans lesquelles elle peut être mise en œuvre pour obtenir les précisions demandées.

► **Remarque sur les approximations numériques**

La condition $\{n \geq 30$ et $np \geq 5$ et $n(1-p) \geq 5\}$, que l'on rencontre souvent dans les documents, est parfois insuffisante, comme le montrent ces deux exemples, pour des valeurs de n et p la remplissant très largement et même dans le cas d'une distribution parfaitement symétrique. Ces approximations étaient systématiquement utilisées à l'époque où l'on ne disposait pas de moyens de calcul facilement accessibles et que de tables partielles des distributions binomiales. Actuellement, elles sont complètement anachroniques dans leur utilisation pour faire des approximations numériques au lycée.
Les lignes de commandes en langage R permettant d'illustrer ceci sont **les lignes 6 à 23 du fichier texte RemarkBac2013.R**, téléchargeable à l'adresse http://www.apmep.asso.fr/-Supplements-en-ligne-au-BV-
Les numéros de ligne ne font pas partie du texte, ils apparaissent dans les éditeurs de codes comme RStudio, Emacs, Tinn-R ou tout autre éditeur de code. C'est ce fichier que vous pouvez utiliser directement dans un éditeur de code ou dont vous pouvez copier-coller les lignes dans la console R, pour les exécuter.
Le fichier **RemarkBac2013CommentaiR.pdf**, téléchargeable à la même adresse, contient toutes les lignes de commande et le code des fonctions **R**, dont il est question dans cet article, avec leurs commentaires.
La correction de continuité est une pratique (elle a ses détracteurs car ce n'est pas un théorème, il n'y a donc pas de démonstration) dont on peut trouver un certain

éclairage dans l'illustration du théorème de Moivre-Laplace. En effet, pour passer d'une variable discrète à une variable continue, on passe de valeurs entières k à des intervalles du type $[k - 0{,}5 \; ; \; k + 0{,}5]$.

Le lecteur désireux d'approfondir cette illustration trouvera la séquence complète, mise en œuvre avec **R**, pages 4 à 6 (activité 1, atelier p2-13) du document téléchargeable sur la page des journées nationales de Marseille sur le site de l'APMEP
(http://www.apmep.asso.fr/IMG/pdf/Atelier_JN_Marseille_13_P2_13_Article.pdf).

B – Distinguer la variable aléatoire F, proportion de succès dans un schéma de Bernoulli, de la valeur *f* prise par cette variable lors d'une réalisation de ce schéma ; parler de prendre une décision par rapport à une proportion dans une population.

► Antilles-Guyane, exercice 2 partie A : L'énoncé confond F et *f*.

PARTIE A

Soient n un entier naturel, p un nombre réel compris entre 0 et 1 et X_n une variable aléatoire suivant une loi binomiale de paramètres n et p.

On note $F_n = \dfrac{X_n}{n}$ et f une valeur prise par F_n.

On rappelle que, pour n assez grand, l'intervalle $\left[p - \dfrac{1}{\sqrt{n}} \; ; \; p + \dfrac{1}{\sqrt{n}}\right]$ contient la fréquence f avec une probabilité au moins égale à 0,95.

En déduire que l'intervalle $\left[f - \dfrac{1}{\sqrt{n}} \; ; \; f + \dfrac{1}{\sqrt{n}}\right]$ contient p avec une probabilité au moins égale à 0,95.

Une rédaction correcte pourrait être :

On rappelle qu'il existe un entier n_0 à partir duquel la variable aléatoire F_n prend ses valeurs dans l'intervalle $\left[p - \dfrac{1}{\sqrt{(n)}} \; ; \; p - \dfrac{1}{\sqrt{(n)}}\right]$ *avec une probabilité au moins égale à 0,95.*

En déduire que, pour n suffisamment grand, l'intervalle $J_n = \left[F_n - \dfrac{1}{\sqrt{(n)}} \; ; \; F_n + \dfrac{1}{\sqrt{(n)}}\right]$ *dont les bornes sont des variables aléatoires, contient p avec une probabilité au moins égale à 0,95.*

► Amérique du nord, exercice 3 Partie B 1. :

Partie B

Les méthodes de production ont été modifiées dans le but d'obtenir 96 % de pains commercialisables.
Afin d'évaluer l'efficacité de ces modifications, on effectue un contrôle qualité sur un échantillon de 300 pains fabriqués.

1. Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % de la proportion de pains commercialisables dans un échantillon de taille 300.
2. Parmi les 300 pains de l'échantillon, 283 sont commercialisables.
   Au regard de l'intervalle de fluctuation obtenu à la question 1, peut-on décider que l'objectif a été atteint ?

• Il serait préférable de distinguer, dans la rédaction, la valeur *f* de la proportion de pains commercialisables observée dans un échantillon, de la variable aléatoire F dont *f* est une réalisation. La mention « proportion », seule dans la phrase, favorise la confusion.

Une rédaction plus précise pourrait être :

> ***1.** Déterminer l'intervalle de fluctuation asymptotique au seuil de 95% de la variable aléatoire F prenant pour valeur la proportion de pains commercialisables dans un échantillon de taille 300.*
>
> ***2.** Parmi les 300 pains de l'échantillon aléatoire, on en a trouvé 283 de commercialisables. En utilisant le résultat trouvé à la question **1.**, peut-on considérer que la proportion de pains commercialisables, obtenue par les méthodes modifiées, est de 96 % ?*

La rédaction des sujets gagnerait à faire clairement la distinction entre valeur observée et variable aléatoire dont elle est une réalisation. Il en est de même concernant les questions sur l'utilisation de l'intervalle de fluctuation pour tester une hypothèse sur une proportion.

► Remarque sur la probabilité binomiale des intervalles de fluctuation asymptotiques :

Le théorème énoncé dans la partie A de l'exercice 2 du sujet Antilles-Guyane porte sur l'intervalle de fluctuation asymptotique du programme de Seconde(1). Je pensais, naïvement, que cette propriété était aussi valable pour l'intervalle de fluctuation asymptotique du programme de Terminale.

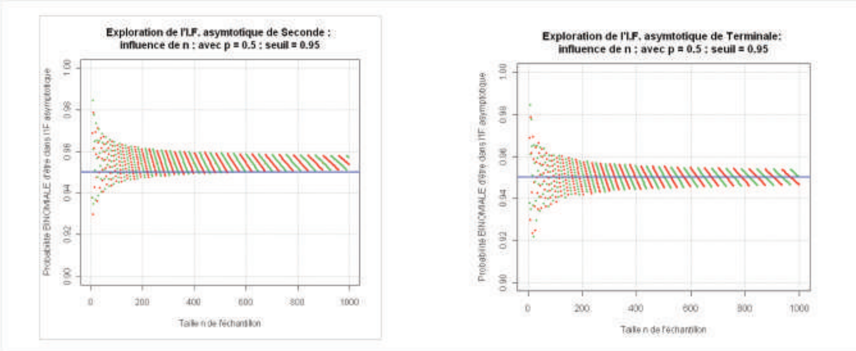
Or on se rend compte sur le graphique de droite ci-après qu'il n'est rien.

C'est à l'occasion de la mise en œuvre d'algorithmes permettant de calculer la probabilité binomiale de ces intervalles de fluctuation asymptotiques, que je me suis rendu compte que ça ne se passait pas de la même façon pour celui de Terminale(2). Aussi loin que j'aie pu aller pour les valeurs de n (j'ai été jusqu'à 105), il y en a toujours pour lesquelles la probabilité binomiale de l'intervalle de fluctuation asymptotique de Terminale (au seuil de 95%) est inférieure à 0,95. Les graphiques suivants permettent de comparer les deux méthodes : le principe de l'algorithme consiste à passer des valeurs de la variable fréquence aux valeurs correspondantes de la variable binomiale X (« discrétisation ») et à calculer la probabilité binomiale de l'intervalle ainsi obtenu. C'est cette "discrétisation" qui explique la forme particulière du nuage de points. L'alternance des couleurs noire et verte correspond à la parité de n.

---

(1) Sous les conditions $n \geq 30$ et $np \geq 5$ et $n(1 - p) \geq 5$, la probabilité de l'intervalle de rayon $\dfrac{1}{\sqrt{n}}$ est voisine de 0,95 mais, contrairement à ce qu'affirme le programme de Seconde, elle n'est pas toujours supérieure à 0,95. On le voit sur le graphique de gauche. On pourra lire à ce propos l'article «Intervalle : de confiance ?» de Louis-Marie BONNEVAL dans le Bulletin n° 427 (NDLR).

(2) L'intervalle de Seconde a pour rayon $\dfrac{1}{\sqrt{n}}$, celui de Terminale $\dfrac{1{,}96}{2\sqrt{n}}$. Ce dernier, étant un peu plus petit, a une probabilité un peu plus faible, donc plus souvent inférieure à 0,95 (NDLR).



Les lignes de commandes en langage **R** permettant de réaliser ces illustrations sont les lignes 56 à 100 du fichier texte **RemarkBac2013.R**. Leurs commentaires figurent dans le fichier **RemarkBac2013CommentaiR.pdf**.

Pour obtenir celui de Seconde, la transformation est courte et facilement réalisable. Il est intéressant de faire réfléchir les élèves sur la raison pour laquelle le seuil de probabilité de 95% ne figure plus dans la partie calcul de la fonction.

Vous retrouverez plus de détails sur ces exemples, ainsi que d'autres, dans le document de l'atelier P2-13 des journées nationales de Marseille, (activité 3 pages 10 à 14,
http://www.apmep.asso.fr/IMG/pdf/Atelier_JN_Marseille_13_P2_13_Article.pdf).

C – Remarques sur les tables de valeurs numériques

Concernant le calcul de probabilités, nous regrettons les questions portant sur la simple lecture de tables de valeurs (ce qui est complètement anachronique !) alors que l'utilisation des calculatrices est un leitmotiv de tous les programmes.

► **Liban exercice 2 partie B question 1 :**

Partie B

1. On note X la variable aléatoire qui, à un petit pot pris au hasard dans la production de la chaîne F_1, associe sa teneur en sucre.
On suppose que X suit la loi normale d'espérance $m_1 = 0,17$ et d'écart-type $\sigma_1 = 0,006$.
Dans la suite, on pourra utiliser le tableau ci-dessous.

α	β	$P(\alpha \leqslant X \leqslant \beta)$
0,13	0,15	0,000 4
0,14	0,16	0,047 8
0,15	0,17	0,499 6
0,16	0,18	0,904 4
0,17	0,19	0,499 6
0,18	0,20	0,047 8
0,19	0,21	0,000 4

Donner une valeur approchée à 10^{-4} près de la probabilité qu'un petit pot prélevé au hasard dans la production de la chaîne F_1 soit conforme.

• C'est une simple lecture de table, même pas « standard ». Quel intérêt, alors que les nouveaux programmes insistent sur l'autonomie que l'on doit donner aux élèves dans l'utilisation des calculatrices ?

Dans la partie A, on a comme donnée « 95 % de petits pots conformes dans la chaîne F1 ». Dans la partie B question 1, pour ce même événement « concret », on trouve $P(16 \leq X \leq 18) \approx 0{,}9044$. Certainement déstabilisant pour les élèves les plus attentifs !

► **Quelques activités alliant algorithmique et calcul de probabilités**

Les calculs de probabilités constituent un excellent terrain pour la conception d'algorithmes et leur mise en œuvre avec les calculatrices programmables. C'est indispensable lorsque les distributions étudiées ne sont pas préprogrammées, comme par exemple la loi géométrique tronquée qui apparaît en *Première S*. Mais même lorsque qu'elles y sont implantées, il est souvent intéressant de construire des programmes mieux adaptés à certaines pratiques.

En effet, même lorsque des distributions préprogrammées existent, il peut être intéressant, voire utile de rechercher et programmer des algorithmes plus pratiques, plus directement utilisables (comme par exemple pour calculer directement $P(A \leq X \leq B)$, ce qui n'est pas fait avec les fonctions préprogrammées), plus adaptés aux définitions particulières du programme comme par exemple celles du mode de détermination de l'intervalle de fluctuation d'une variable binomiale ou d'une variable proportion.

Je propose tous les ans à mes élèves un document contenant quelques programmes de calcul de probabilités à implanter dans leur machine, après en avoir fait une analyse détaillée permettant de comprendre leur fonctionnement et leur utilité. Ce document (**ProgCalculettesBinoCommente.pdf**) est disponible à l'adresse http://revue.sesamath.net/spip.php?article552 .

X étant une variable aléatoire de loi binomiale de paramètres N et P (majuscules par souci d'homogénéité de notation pour la programmation sur la majorité des calculettes), on y trouve :

• Calcul de $P(A \leq X \leq B)$ pour $0 \leq A \leq B \leq N$. Particulièrement utile pour les élèves ayant des vieilles versions de calculettes ne possédant pas la distribution binomiale préprogrammée.

• Calcul de toutes les valeurs de la distribution et de la répartition de X, (pour N de taille dépendant de la capacité mémoire de la calculette), les résultats étant stockés dans des listes. Permet de voir l'utilisation des listes et de certaines fonctions préprogrammées relatives aux listes et aux distributions binomiales.

• Calcul de l'Intervalle de Fluctuation bilatéral de X, selon le mode de calcul du document d'accompagnement de *Première S*. On utilise l'algorithme précédent pour implémenter une liste avec les valeurs de la répartition de X. Une boucle « Pour » parcourt la liste et détermine les valeurs A et B de l'I.F. [A ; B]. Le programme Texas utilise la répartition binomiale préprogrammée disponibles dès la TI82. Le programme Casio n'utilise que la fonction seq(...), disponible dès la Casio Graph 25.

• Calcul des Intervalles de Fluctuation bilatéraux de X et de X / N, selon le mode de calcul du document d'accompagnement de *Première S*. Les programmes utilisent directement une boucle « TantQue » et n'utilisent pas les listes.

Le paragraphe suivant va être l'occasion de présenter quelques activités d'algorithmiques sur le calcul de distributions de probabilités, mises en œuvre avec **R**.

D – Remarques sur l'algorithmique :

• Les premiers exercices sont de facture classique, avec cependant des niveaux de difficulté très variés : proposer une modification ou un algorithme (sujet Métropole ...) est plus difficile que calculer des valeurs résultant d'un algorithme (sujet Antilles-Guyane …).

• L'exercice 4 du sujet 2013 de septembre d'Antilles-Guyane propose un algorithme de simulation d'une marche aléatoire. Alors que l'énoncé précise : « L'objectif de cet exercice est d'estimer la probabilité p de l'événement S [le robot] Tom traverse le pont », on constate que ni l'algorithme de la partie A, ni le calcul de la partie B ne proposent une estimation de cette probabilité. La partie A propose la simulation d'une marche aléatoire. Or il faudrait faire tourner un grand nombre de fois cet algorithme pour avoir une estimation de p. C'est ce que je propose dans un prolongement. La partie B permet de trouver une valeur approchée de cette probabilité, en utilisant des suites.

Exercice 4 **5 points**
Candidats n'ayant pas suivi l'enseignement de spécialité

Les deux parties sont indépendantes

Le robot Tom doit emprunter un pont sans garde-corps de 10 pas de long et de 2 pas de large. Sa démarche est très particulière :

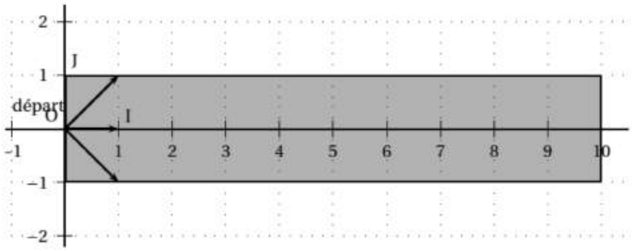
- Soit il avance d'un pas tout droit ;
- Soit il se déplace en diagonale vers la gauche (déplacement équivalent à un pas vers la gauche et un pas tout droit) ;
- Soit il se déplace en diagonale vers la droite (déplacement équivalent à un pas vers la droite et un pas tout droit).

On suppose que ces trois types de déplacement sont aléatoires et équiprobables.

L'objectif de cet exercice est d'estimer la probabilité p de l'évènement S « Tom traverse le pont » c'est-à-dire « Tom n'est pas tombé dans l'eau et se trouve encore sur le pont au bout de 10 déplacements ».

Partie A : modélisation et simulation

On schématise le pont par un rectangle dans le plan muni d'un repère orthonormé (O , I, J) comme l'indique la figure ci-dessous. On suppose que Tom se trouve au point de coordonnées $(0\,;0)$ au début de la traversée. On note $(x\,;\,y)$ les coordonnées de la position de Tom après x déplacements.

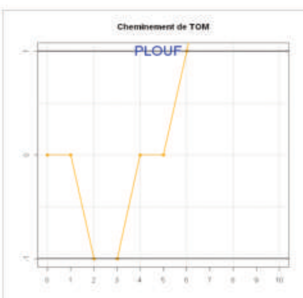


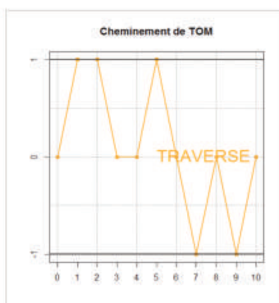
On a écrit l'algorithme suivant qui simule la position de Tom au bout de x déplacements :

```
x, y, n sont des entiers
Affecter à x la valeur 0
Affecter à y la valeur 0
Tant que y ⩾ −1 et y ⩽ 1 et x ⩽ 9
        Affecter à n une valeur choisie au hasard entre −1, 0 et 1
        Affecter à y la valeur y + n
        Affecter à x la valeur x + 1
Fin tant que
Afficher « la position de Tom est » (x ; y)
```

1. On donne les couples suivants : $(-1\,;\,1)$; $(10\,;0)$; $(2\,;4)$; $(10\,;2)$.
   Lesquels ont pu être obtenus avec cet algorithme ? Justifier la réponse.
2. Modifier cet algorithme pour qu'à la place de « la position de Tom est $(x\,;\,y)$ », il affiche finalement « Tom a réussi la traversée » ou « Tom est tombé ».

• La traduction en langage **R** de l'algorithme de l'énoncé suivi d'un prolongement permettant une représentation graphique du trajet du robot et d'un prolongement permettant de calculer une estimation de la probabilité de S, par simulation, figurent aux lignes 152 à 262 du fichier texte **RemarkBac2013.R**. Leurs commentaires figurent dans le fichier **RemarkBac2013CommentaiR.pdf**.





E – Les limites du « langage naturel » et l'utilisation des listes, en algorithmique

• L'APMEP et la C2I tiennent à faire remarquer que le « langage naturel » préconisé pour l'écriture des algorithmes est beaucoup trop imprécis pour donner lieu à l'évaluation d'algorithmes plus élaborés. Il serait alors nécessaire d'utiliser les langages de programmation habituels. C'est, par exemple, dans la description des opérations sur les listes que le « langage naturel » montre ses limites.

• Il est dommage qu'aucun sujet ne propose l'utilisation de listes alors que ce sont des objets couramment utilisés en programmation moderne.

L'exemple choisi porte sur le calcul d'un intervalle de fluctuation d'une variable binomiale X dont on déduit directement l'intervalle de fluctuation de la variable fréquence correspondante X/n. Je vais comparer deux algorithmes, l'un classique utilisant la boucle « tant que », l'autre appliquant directement la définition à des listes.

L'intervalle de fluctuation bilatéral au seuil minimal de probabilité s d'une variable binomiale X de paramètres n et p, défini dans le document ressource du programme de Première S, est l'intervalle $[a\ ;\ b]$ où
a est la plus petite valeur de X telle que $P(X \leq a) > (1 - s) / 2$
b est la plus petite valeur de X telle que $P(X \leq b) \geq 1 - (1 - s) / 2$
$[a/n\ ;\ b/n]$ est l'intervalle de fluctuation de la variable fréquence (de succès) X/n.

Algorithme faisant intervenir les opérations sur les listes :

Le langage **R** permet de traduire directement la définition de *première S*. Dans la fonction PIF_Bino(...) :
La ligne repartX <- pbinom(0:n, n, p) implémente le vecteur repartX avec la répartition de X.
La ligne rang_a <- min(which(repartX > (1 - s) / 2)) prend le plus petit des rangs (pas de rang 0) des valeurs de repartX strictement supérieures à (1 - s) / 2).
La ligne a <- rang_a - 1 retrouve la valeur de X à partir de son rang (décalage de 1).

Les lignes de commandes en langage **R** mettant en œuvre cette définition, à l'aide des opérations sur les listes sont les lignes 106 à 118 du fichier texte **RemarkBac2013.R**
(commentées dans le fichier **RemarkBac2013CommentaiR.pdf**).

La seule difficulté de l'algorithme réside dans la bonne mise en œuvre, donc la compréhension, du mode de calcul proposé. On évite ainsi de superposer la difficulté de l'élaboration de l'algorithme, à la difficulté de compréhension du mode de calcul.

L'algorithme classique suivant est tiré de l'exercice 11 page 331 du manuel de Première S Nathan Hyperbole. Il fait intervenir la boucle tant que :

Algorithmique

11 Algorithme et intervalle de fluctuation

OBJECTIF Déterminer l'intervalle de fluctuation à l'aide d'un algorithme.

X est une variable aléatoire qui suit une loi binomiale de paramètres *n* et *p*.

1. Détermination de *a*

On étudie l'algorithme suivant :

```
Entrées
Saisir n et p
Initialisations
e prend la valeur 0,05
k prend la valeur 0
S prend la valeur 0
Traitement
Tant que S ≤ e/2
    | S prend la valeur S + P(X = k)
    | k prend la valeur k + 1
FinTantque
a prend la valeur k − 1
Sortie
Afficher a
```

a) Faire fonctionner cet algorithme à la main dans le cas particulier où $n = 10$ et $p = 0{,}7$.
Pour cela, compléter le tableau ci-dessous en indiquant les valeurs successives des variables *k* et S.

S	0	$5{,}9049 \times 10^{-6}$				
k	0	1				

On pourra utiliser les données suivantes :

k	P(X<=k)
0	5,9049E-06
1	0,000143686
2	0,001590386
3	0,010592078
4	0,047348987
5	0,150268333
6	0,350389282
7	0,617217214
8	0,850691654
9	0,971752475
10	1

Quelle est la valeur de *a* affichée dans ce cas ?

b) Expliquer le rôle de cet algorithme.
Préciser la signification de chacune des variables.

c) Coder cet algorithme dans un langage de l'ordinateur ou de la calculatrice.
Voici par exemple, le programme obtenu dans le langage Scilab.

```
1  n=input("n=");
2  p=input("p=");
3  e=0.05
4  k=0
5  S=0;
6  while S<=e/2
7    S=S+[illegible](n,k)*p^k*(1-p)^(n-k);
8    k=k+1;
9  end
10 a=k-1
11 afficher("a="+string(a))
```

Note : *Dans une prochaine version, le langage Scilab possédera une fonction calculant directement $P(X = k)$.*
La ligne 7 s'écrira alors :
$S = S + loi\_binomiale(n, p, k)$.

d) Tester le programme obtenu avec différentes valeurs de *n* et *p*.

2. Détermination de *b*

a) Écrire de même l'algorithme qui détermine le plus petit entier *b* tel que :

$$P(X \leqslant b) \geqslant 0{,}975$$

b) Coder cet algorithme dans un langage de programmation.

c) Tester le programme obtenu.

3. Varier le seuil de décision

a) Modifier les programmes précédents afin que le seuil de décision soit fixé à 85 %.

b) Tester les corrections apportées aux programmes.

Les lignes de commandes en langage **R** mettant en œuvre cette définition, à l'aide des opérations sur les listes sont les lignes 124 à 146 du fichier texte **RemarkBac2013.R**,
commentées dans le fichier **RemarkBac2013CommentaiR.pdf**.

Cet algorithme superpose les difficultés suivantes :

• La première et principale difficulté est la bonne gestion de la boucle TantQue : la plus petite valeur … telle que $P(X \leq a) > e/2$ se traduit par TantQue(S <= e / 2). Il faut inverser le sens de l'inégalité.

• La seconde est la place du compteur k, pour trouver la valeur de a.

• La troisième est le calcul des probabilités binomiales et leur cumul, que l'on contourne en utilisant dbinom(...) et pbinom(...)

• 15 lignes (hors les affichages) contre 7 lignes pour la version « liste ».

Nous venons de toucher du doigt les difficultés engendrées par le fait d'utiliser le « langage naturel », ce qui limite les possibilités d'utiliser des fonctions spécifiques aux « vrais » langages de programmation et qui rend les algorithmes plus complexes.

Pour que les élèves s'investissent en algorithmique, il faut qu'ils puissent mettre en œuvre facilement de véritables outils d'illustration du cours et de résolution de problème, ce que seule permettra l'utilisation des langages modernes de programmation.