Bases d'exercices à destination des enseignants du secondaire

Intervalles de confiance - Intervalles de fluctuation - Prises de décision

IREM Bordeaux - groupe Probabilités et Statistique - 2013

... ces exercices ne sont pas forcément à destination des élèves

Exercice 1. On dispose d'un médicament efficace dans le traitement d'une maladie dans 80% des cas. Une variante de ce médicament est testée, et on observe 32 guérisons sur 50 patients. Peut-on considérer que cette variante a aussi une efficacité de 80% au seuil 95 % ?

La question est formulée pour être traitée dans le cadre du programme de lycée avec un intervalle de fluctuation asymptotique au seuil 95% *pour une proportion. On devrait d'ailleurs décomposer le problème avec une première question portant sur le calcul de l'intervalle de fluctuation au seuil* 95%*.*
Supposons que la variante a la même efficacité. L'utilisation d'un intervalle de fluctuation asymptotique pour une proportion donne avec les formules de terminale $[0.69; 0.91]$*. On observe une proportion de 32/50=0.64, qui n'appartient pas à l'intervalle de fluctuation. Les données de l'échantillon permettent de conclure que la variante n'a pas une efficacité de* 80%*, ceci avec un risque de* 5% *de se tromper.*
Remarque sur la question des arrondis dans les intervalles de fluctuation : *Pour garantir la probabilité d'un intervalle de fluctuation* $[a_n, b_n]$*, on peut proposer d'arrondir par défaut la valeur de* a_n *et par excès la valeur de* b_n*.*

Exercice 2. On dispose d'un médicament efficace dans le traitement d'une maladie dans 80% des cas. Une variante de ce médicament est testée, et on observe 35 guérisons sur 50 patients. Peut-on supposer que cette variante a aussi une efficacité de 80% au seuil 95 % ?

La question est formulée pour être traitée dans le cadre du programme de lycée avec un intervalle de fluctuation asymptotique pour une proportion. Supposons que la variante a la même efficacité. L'utilisation d'un intervalle de fluctuation asymptotique pour une proportion donne avec les formules de terminale $[0.69; 0.91]$*. On observe une proportion de 35/50=0.70, qui appartient à l'intervalle de fluctuation. On conclut donc que les données de l'échantillon ne contredisent pas la supposition que la variante peut être efficace avec un taux de 80%. On évite pourtant de dire qu'on accepte la supposition car on ne maîtrise pas le risque d'erreur dans ce cas.*

Exercice 3. On dispose d'un médicament efficace dans le traitement d'une maladie dans 80% des cas. Une variante de ce médicament est testée, et on observe 35 guérisons sur 50 patients. Peut-on supposer que cette variante a la même efficacité que le médicament d'origine au seuil 95 % ?

On peut observer que l'énoncé est trés semblable au précédent et se résout donc de la même façon si on comprend le terme " même efficacité" comme "efficacité = 80%". Si on comprend le mot même comme comparable, on peut penser à utiliser un intervalle de confiance. Les énoncés doivent donc être précis.

Exercice 4. Un fabriquant d'objets annonce que 10% de sa production est défectueuse. Un statisticien prend un lot de 60 pièces. Sur ces 60, 8 sont défecteux.
Est-ce cohérent à un seuil de 95% avec l'annonce faite ?
Si le statisticien choisit un lot de 20 pièces, et trouve 3 défectueux, peut-on appliquer la même procédure ?

(peut se faire en première) : La question est un problème de prise de décision à l'aide d'un intervalle de fluctuation. La taille de l'échantillon (60) est assez grande pour pouvoir faire de l'asymptotique et on est bien dans le cadre du programme de terminale.
Le nombre d'objets défectueux dans l'échantillon suit une loi binomiale de paramètres $(60, 0.1)$. *On trouve l'intervalle de fluctuation au seuil 0.95 pour la proportion dans un échantillon de 60 pièces de* $[0.024; 0.18]$.
La proportion observée est de 8/60=0.13, elle est bien dans l'intervalle de fluctuation donc les données ne permettent pas de remettre en cause l'annonce faite.
La deuxième question soulève le cas d'un échantillon hors contexte du programme de terminale, car trop petit (n=20) pour pouvoir appliquer l'intervalle de fluctuation asymptotique. On utilise donc le calcul d'un intervalle de fluctuation basé sur la loi binomiale (programme de première).

Exercice 5. En septembre 2005, 13% des Talençais prennent la ligne B au moins une fois par semaine
En février 2006, une enquête a été effectuée auprès de 200 Talençais. 36 d'entre eux ont déclaré prendre la ligne B au moins une fois par semaine.
Peut-on dire au seuil 95% que la plus grande fiabilité de la ligne a entrainé une hausse de sa fréquentation par les Talençais ? Même question avec un seuil 99% ?

Si 13% *des Talençais prennent encore le tram au moins une fois par semaine en février 2006, alors un intervalle de fluctuation au seuil 95% pour la proportion sur un échantillon de 200 personnes est* $[0.083; 0.177]$.
Puisque la fréquence observée, qui vaut 36/200=0.18, est hors de l'intervalle, on peut considérer qu'il y a eu un changement dans la fréquentation, ceci au risque de 5% *de se tromper mais rien ne permet de relier ceci à la hausse de la fiabilité.*
Au seuil de 99%, l'intervalle donne $[0.069; 0.191]$ *et la conclusion change : on ne peut rien conclure pour ce seuil. Les données ne permettent pas de conclure qu'il y a eu une modification de la fréquentation.*

Exercice 6. Deux enquêtes ont été effectuées auprès de 200 Talençais en septembre 2005 et février 2006. En septembre 13 % des personnes prenaient la ligne B au moins une fois par semaine, leur nombre s'élève à 36 sur 200 en février. Dans le même temps le nombre de pannes survenues sur la ligne a diminué de 26%. Les données permettent-elles d'appliquer la technique de prise de décision (par intervalle de fluctuation) ?

Non : ici, les deux pourcentages sont des observations sur des échantillons. Un intervalle de fluctuation pour une proportion nécessite de connaître la vraie probabilité pour une loi Binomiale. Ici on a deux proportions observées. L'exercice est repris plus loin dans le cadre des intervalles de confiance.

Exercice 7. Un laboratoire annonce que l'un de ses médicaments est efficace à 95%. On teste cette affirmation, et on obtient que sur un échantillon de 400 personnes traitées, le traitement a été efficace pour 369 d'entre elles. Peut-on accepter l'affirmation du laboratoire au seuil 0.99 ? Au seuil 0.9 ?
A partir de quel seuil commence-t-on à rejeter l'affirmation ?

Il s'agit là encore de décision : le terme "accepter" dans le texte est maladroit (cf exercice 2), l'énoncé est équivalent dans sa résolution si on le remplace par "rejeter". On cherche à savoir si les observations sont en accord avec une proportion de guérison pour la population totale (des gens traités) de 0.95. *(L'une des difficultés ici peut être cette valeur de* 0.95 *qui ici ne représente pas le seuil de décision du test...).*
On peut faire le calcul en non-asymptotique avec la loi binomiale $B(400, 0.95)$ *mais pas au seuil de* 0.99 *ce qui sort du cadre de la classe de 1ère. On peut aussi se placer en asymptotique. La recherche du seuil auquel on rejette l'affirmation peut elle aussi se faire aussi bien en asymptotique qu'en non asymptotique.*

Dans un cadre asymptotique, si on suppose une probabilité de guérison de 0.95*, un intervalle de fluctuation asymptotique pour la fréquence dans un échantillon de taille 400 est, au seuil 0.99,* $[0.95 - 2.58\sqrt{\frac{0.95\times0.05}{400}}; 0.95 + 2.58\sqrt{\frac{0.95\times0.05}{400}}] = [0.922; 0.979]$. *369/400 appartient à cet intervalle, on ne peut donc pas rejeter l'affirmation du laboratoire. Au seuil 0.9, on remplace* 2.58 *par* 1.65 *et on obtient l'intervalle* $[0.932; 0.968]$. *Contrairement au cas précédent, ce seuil permet de rejeter l'affirmation, ceci au risque 10% de se tromper.*
On change de décision lorsque la borne inférieure de l'intervalle vaut 0.9225*, c'est-à-dire pour* $u_\alpha = 2.524$. *Or l'utilisation de la calculatrice permet d'obtenir, si* X *suit une loi normale centrée réduite,* $P(-2.524 \leq X \leq 2.524) = 0.988$. *C'est donc le seuil critique de changement de décision.*

Exercice 8. On dispose d'un médicament efficace dans le traitement d'une maladie dans 80% des cas. Une variante de ce médicament est testée, et on observe 43 guérisons sur 50 patients. L'efficacité de cette variante est-elle différente de celle du médicament d'origine au seuil 95 % ? Jusqu'à quel seuil peut-on dire que la variante a eu un effet ?

La première question est un problème de décision classique mais peu guidé. La question est formulée pour être traitée dans le cadre du programme de lycée (intervalle de fluctuation symétrique), mais on pourrait aussi se demander si la variante est "significativement plus efficace" ce qui conduirait à répondre par un intervalle unilatéral. La deuxième question est moins standard et consiste à calculer ce qu'on appelle une p-valeur : on considère le seuil $1-\alpha$ *(qui valait 0.95) comme un paramètre. Si cette valeur change, la décision avec la même observation, peut changer. La p-valeur est la valeur de* α *à laquelle on change de décision.*
Pour la première question, une difficulté est de bien comprendre que 80% *correspond à une probabilité théorique donnée. Il s'agit de voir si la fréquence observée* $43/50$ *est compatible avec cette probabilité. Il s'agit donc d'un problème de décision, qui au niveau lycée s'aborde à l'aide d'intervalle de fluctuation. Si on suppose que la variante a la même efficacité et guérit* 80% *des malades, le nombre* N *de guérisons sur* 50 *patients suit une loi binomiale de paramètres* $(50, 0.8)$*. On obtient alors à l'aide d'un tableur un intervalle de fluctuation au seuil de* 95% $[34; 45]$ *pour* N*, soit un intervalle de fluctuation pour la proportion de guérison dans un échantillon de 50 patients de* $[0.68; 0.90]$*. La proportion observée est de* 0.86 *: elle appartient à l'intervalle de fluctuation ; donc au seuil de* 95% *on ne peut pas dire que la variante ait une efficacité significativement différente au vu des observations.*
Remarques sur cette première question : l'utilisation d'un intervalle de fluctuation aurait donné avec les formules de 2nde $[0.8-\frac{1}{\sqrt{50}}, 0.8+\frac{1}{\sqrt{50}}] = [0.66; 0.94]$ *(intervalle plus gros) et avec les formules de terminale également asymptotique* $[0.69; 0.91]$*, très proche de l'intervalle obtenu en non asymptotique : on est bien ici dans le cadre d'une binomiale où la taille de la population est importante.*
Si la question avait été « Est-ce que la variante a une efficacité significativement supérieure », il aurait pu être plus naturel de calculer un intervalle de fluctuation unilatéral, de la forme $[0, \frac{b}{50}]$*, et de chercher le nombre* b *tel qu'une binomiale de paramètre* $(50, 0.8)$ *est strictement supérieure à* b *avec une probabilité inférieure à* 0.05 *(cf le cas de Woburn dans le document d'accompagnement de 1ère) : on aurait alors trouvé* $b = 44$ *soit un intervalle de fluctuation pour la proportion de guérison dans un échantillon de 50 malades de* $[0, 0.88]$*.*

On aurait décidé que la variante était plus efficace si la borne supérieure de l'intervalle de fluctuation pour $\frac{N}{50}$ *avait été* 0.86 *au lieu de* 0.90*. La probabilité qu'une binomiale de paramètres* $(50, 0.8)$ *soit strictement supérieure à* 42 *est de* 0.19*. Dans le cadre bilatéral du programme, il aurait pour cela fallu calculer un intervalle de fluctuation au seuil de* 2×0.19 *soit* $\alpha = 0.38$*. Dans un cadre unilatéral cela aurait correspondu à* $\alpha = 0.19$*. C'est une valeur assez élevée, signe que cette différence (*0.86 *par rapport à* 0.8*) n'est pas très significative sur un échantillon aussi petit.*

Exercice 9. Dans une commune, en 2001, on a enregistré 418 naissances de garçons et 388 naissances de filles. Le maire annonce "Dans ma commune, il naît autant de filles que de garçons."
Dans tout le pays, en 2001, 41800 garçons et 38800 filles sont nés. Le président annonce "Dans mon pays, il naît autant de filles que de garçons."
L'un des deux élus a une formation en statistiques, peut-on deviner lequel ?

Exercice peu guidé. Ce qui est suggéré est de voir si les observations sont en accord avec l'affirmation "il naît autant de filles que de garçons". Bien évidemment dans les deux cas, l'affirmation est rigoureusement fausse : les nombres de naissances de filles et de garçons sont différents. Il faut voir si ces deux affirmations peuvent être juste en un sens statistique : l'affirmation serait à prendre au sens "la probabilité qu'un enfant soit une fille est de 0.5" ou encore "la proportion des naissances de filles est 0.5" sous-entendu dans une population totale (fictive), que l'on peut comprendre comme l'ensemble de toutes les naissances sur toutes les années passées et à venir : l'année 2001 correspondant alors à un échantillon.
On peut sans faire les calculs deviner que c'est le maire de la petite commune qui peut avoir raison : plus l'échantillon est petit plus on peut s'attendre à ce qu'il y ait un écart entre la fréquence observé (0.4814) et la fréquence théorique (0.5). Une fois le cadre précédent posé il est facile de faire les calculs, en asymptotique ou pas : un intervalle de fluctuation pour la fréquence correspondant à l'observation d'une binomiale de paramètre $(806, 0.5)$ *est au seuil 0.95* $[0.465, 0.53]$ *: 388/806 est dans cet intervalle, on ne peut pas rejeter l'affirmation du maire.*
Par contre, si on suppose toujours $p = 0.5$*, un intervalle de fluctuation asymptotique pour la proportion de filles dans un échantillon de 80600 naissances est* $[0.4965; 0.5035]$*. La fréquence observée de 0.4814 n'est pas dans l'intervalle, on doit donc rejeter l'affirmation du président, qui n'est probablement pas statisticien.*

Exercice 10. On mesure le poids de 100 paquets de café. On trouve que 8 d'entre eux pèsent moins de 250 grammes.

(1) Le fabricant de café déclare que 7% de ses paquets de café contiennent en fait moins de 250 grammes. Les observations sont-elles compatibles au seuil 95% avec cette affirmation ?

(2) Une association de consommateurs affirme que ce sont en fait 10% des paquets qui pèsent moins de 250 grammes. Les observations sont-elles compatibles au seuil 95% avec cette deuxième affirmation ?

(3) Combien faudrait-il peser de paquets de café pour être sûr de rejeter au seuil de 95% au moins l'une des deux affirmations précédentes ?

Les deux premières questions sont des problèmes de décision qui rentrent dans le cadre du programme de 1ère. La dernière question, si on veut la traiter sans un tatonnement un peu pénible, demande à être faite dans un cadre asymptotique avec le programme de terminale (ou de 2nde), qui trouve alors toute son utilité.
Si on suppose $p = 0.07$*, on trouve comme intervalle de fluctuation pour une loi binomiale de paramètre* $(100, 0.07)$ *au seuil* 0.95 $[0.02; 0.12]$. $8/100$ *appartient à l'intervalle, les observations ne conduisent pas à rejeter l'hypothèse* $p = 0.07$*. Si on suppose* $p = 0.1$*, l'intervalle de fluctuation pour une binomiale* $B(100, 0.1)$ *est* $[0.05, 0.16]$ *: là encore on ne rejette pas l'hypothèse* 0.1 *:* 100 *est beaucoup trop petit pour pouvoir discriminer efficacement entre 0.1 et 0.07.*
On sera sûr de rejeter au moins l'une des deux hypothèses si les deux intervalles de fluctuation correspondants sont disjoints : dans un cadre asymptotique, on trouve que les deux intervalles de fluctuation seront disjoints si $0.07 + 1.96 \times \frac{\sqrt{0.07\times0.93}}{\sqrt{n}} \leq 0.1 - 1.96 \times \frac{\sqrt{0.1\times0.9}}{\sqrt{n}}$ *soit* $n \geq 1316$.

Exercice 11. On effectue un sondage auprès de 400 électeurs et on obtient 212 intentions de vote en faveur d'un candidat C.
Donner au risque $\alpha = 0.05$, un intervalle de confiance des intentions de vote en faveur de C dans la population entière.

On trouve $[\frac{212}{400} - \frac{1}{\sqrt{400}}; \frac{212}{400} + \frac{1}{\sqrt{400}}] = [0,48; 0,58]$ *: au niveau* 0.95, *C peut simplement dire que les intentions de vote en sa faveur (sur la population) est une valeur comprise entre 48% et 58%.*

Exercice 12. Un biologiste veut évaluer la proportion p de drosophiles à yeux rouges dans une population de drosophiles. Il prend un échantillon de 80 drosophiles et observe 12 drosophiles aux yeux rouges. Avec une autre population soumise à des conditions différentes, il observe dans un échantillon de même taille une proportion de 0.25.
Donner les intervalles de confiance pour les proportions de drosophiles à yeux rouges dans les deux populations au niveau de confiance 95%.

Avec la formule $[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}]$, *les intervalles sont égaux à* $[0.038; 0.262]$ *et* $[0,138; 0,262]$

Exercice 13. Un biologiste veut évaluer la proportion p de drosophiles à yeux rouges dans une population de drosophiles. Il prend un échantillon de 80 drosophiles et observe 12 drosophiles aux yeux rouges. Avec une autre population soumise à des conditions différentes, il observe dans un échantillon de même taille une proportion de 0.25. Il déclarera que ces conditions ont un effet significatif si les deux intervalles de confiance sont disjoints. Quelle sera sa conclusion ?

Il s'agit ici de comparer deux proportions inconnues, en comparant deux intervalles de confiance ; l'intervalle de confiance est donc utilisé pour un problème de décision, mais dans un cadre où on ne sait *rien sur la proportion* p *; le premier intervalle est, avec la formule* $[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}]$ $[0.038; 0.262]$. *Pour le deuxième intervalle on trouve* $[0.138; 0.362]$. *Ils ne sont pas disjoints et on ne peut pas affirmer que les conditions différentes ont eu un effet statistiquement significatif. A noter que l'autre formule (qui n'est pas au programme)* $[f - 1.96\frac{\sqrt{f(1-f)}}{\sqrt{n}}, f + 1.96\frac{\sqrt{f(1-f)}}{\sqrt{n}}]$ *donne pour le premier intervalle* $[0.072; 0.228]$ *et pour le deuxième* $[0.155; 0.345]$ *: les intervalles sont plus précis (et rétrécis) mais cela ne change rien à la conclusion.*

Exercice 14. On cherche à tester l'efficacité d'un médicament. Pour cela, on prend une population de 100 moutons malades auxquels on fait suivre le traitement, et on observe que 63 d'entre eux sont guéris au bout de deux semaines. Donner au niveau de confiance 0.95 un intervalle de confiance pour la proportion de moutons guéris après un traitement de 2 semaines.
Sur combien de moutons faudrait-il faire le traitement pour obtenir au niveau de confiance 0.95 un intervalle de confiance d'amplitude 0.01 ?
Avec un deuxième traitement sur une autre population de 80 moutons malades, on observe la guérison de 57 moutons. Donner au niveau de confiance 0.95 un intervalle de confiance pour la proportion théorique de guérison avec ce deuxième traitement.
La différence entre les deux traitements est-elle significative ?

Les deux premières questions sont standards. La dernière question sous-entend le même mode de décision que dans des calculs précédents : on dira que la différence est significative si les deux intervalles sont disjoints. On rappelle qu'on ne peut pas ici faire d'intervalle de fluctuation puisqu'on ne connaît pas la vraie valeur...
On trouve pour le premier intervalle $I_1 = [0.53; 0.73]$. *La longueur d'un intervalle de confiance est* $\frac{2}{\sqrt{n}}$ *: pour avoir une longueur de* 0.01 *il faudrait prendre* $n = 40000$, *ce qui n'est pas envisageable pratiquement. Le deuxième intervalle est* $I_2 = [0.60, 0.82]$. *Les deux intervalles ne sont pas disjoints, la différence n'est pas significative, sans qu'on puisse contrôler le risque d'erreur de se tromper*

Exercice 15. Deux enquêtes ont été effectuées auprès de 200 Talençais en septembre 2005 et février 2006. En septembre 13 % des personnes prennent la ligne B au moins une fois par semaine, leur nombre s'élève à 55 sur 200 en février. Dans le même temps le nombre de pannes survenues sur la ligne a diminué de 26%. Peut-on dire au seuil 95% que la plus grande fiabilité de la ligne a entrainé une hausse de sa fréquentation par les Talençais ?

L'énoncé de cet exercice, s'il n'est pas modifié, est stricto senso hors programme de lycée : en effet on est amené à faire de la décision (la proportion de gens prenant le tram a-t-elle augmenté) mais la proportion initiale (de référence) n'est pas connue (le 13% vient également d'un échantillon de 200 personnes). On ne peut donc pas faire de fluctuation. Une méthode pourrait être de calculer deux intervalles de confiance pour ces proportions inconnues, un pour septembre et un pour janvier, et de regarder s'ils s'intersectent. Un tel calcul dans le cadre du programme ne se fait qu'au niveau 95%.
Méthode par comparaison d'intervalle de confiance : soit p_S la proportion des personnes prenant le tram en septembre dans la population totale et p_F la proportion des personnes prenant le tram en février dans la population totale.
Un intervalle de confiance pour p_S au niveau 0.95 est $I_S = [0.13 - \frac{1}{\sqrt{200}}; 0.13 + \frac{1}{\sqrt{200}}] \simeq [0.059; 0.200]$
(l'autre formule mentionnée dans le programme, mais hors programme, $[0.13 - 1.96\sqrt{\frac{0.13\times0.87}{200}}; 0.13 - 1.96\sqrt{\frac{0.13\times0.87}{200}}]$ *fournit* $[0.083; 0.177]$*)*
Pour p_F on obtient :
$I_F = [0.2.75 - \frac{1}{\sqrt{200}}, 0.275 + \frac{1}{\sqrt{200}}] \simeq [0.204; 0.345]$.
Les deux intervalles sont disjoints, on peut dire que la fréquentation a augmenté.

Exercice 16. Une audimétrie a permis d'établir, avec un niveau de confiance de 95%, que l'audience d'une émission télévisée était dans l'intervalle [35%, 45%]. Afin d'affiner la mesure, on décide de sonder des téléspectateurs. Combien faut-il en sonder au minimum pour que, avec le même niveau de confiance de 95%, et, en supposant que l'on obtienne la même estimation ponctuelle de 40% d'audience, on ait un intervalle de confiance de longueur 2% au lieu de 10% ?

Exercice sans aucune difficulté si l'on prend la formule du programme $[f - \frac{1}{\sqrt{n}}, f + \frac{1}{\sqrt{n}}]$*, si ce n'est que l'énoncé fournit beaucoup trop de données : la taille de l'intervalle est* $\frac{2}{\sqrt{n}}$ *et on obtient* $n \geq 10000$*. Les données sont pertinentes si l'on prend la formule (plus précise mais injustifiable en lycée)*
$[f - 1.96\frac{\sqrt{f(1-f)}}{\sqrt{n}}; f - 1.96\frac{\sqrt{f(1-f)}}{\sqrt{n}}]$*. On trouve alors* $n \geq 9220$ *ce qui ne change pas grand chose*

Exercice 17. Un quotidien publie la cote du chef de l'état à partir d'un sondage réalisé auprès de 1000 personnes. Au mois de janvier la cote était de 38% d'opinion favorables, en février 36%. Et le journaliste de commenter ≪ le chef de l'état perd 2 points ≫.
Commenter ce commentaire.

On peut par exemple comparer les deux intervalles de confiance pour la proportion de gens qui ont une opinion favorable dans la population tout entière : on trouve en janvier $[0.35; 0.41]$ *et en février* $[0.33; 0.39]$ *: les deux intervalles ne sont pas disjoints, loin de là. Il est donc tout-à-fait possible que la proportion dans la population entière n'ait pas bougé (voire même qu'elle ait augmenté...) : une variation de 2 points sur un échantillon de 1000 personnes n'est pas significative. On peut d'ailleurs faire cette remarque sans même faire les calculs précédents : la taille de l'intervalle de confiance avec un échantillon de 1000 personnes est* 0.06 *: une variation de 2 points n'est pas significative.*

Exercice 18. On veut construire sur un campus universitaire deux restaurants : le RU1 de N_1 places et le RU2 de N_2 places. On suppose que n étudiants prendront leur repas dans l'un de ces deux restaurants. On prévoit que les étudiants choisiront le RU1 avec une probabilité p et dans le RU2 avec une probabilité $1-p$. Leurs choix sont indépendants. On suppose enfin que $n \geqslant 200$.

(1) On note X_n le nombre d'étudiants choisissant le RU1 pour un repas donné. Déterminer la loi de X_n.

(2) Écrire l'intervalle de fluctuation asymptotique I_n au seuil de $0,95$ de la variable $\frac{X_n}{n}$.

(3) L'administration désire que les n étudiants trouvent une place dans le RU de leur choix avec une probabilité d'au moins $0,95$ à $0,01$ près.

Déterminer l'intervalle J_n dans lequel doit se trouver $\frac{X_n}{n}$ avec une probabilité d'au moins $0,94$.

(4) On suppose que les conditions sont réunies pour que l'on puisse utiliser l'approximation $P\left(\frac{X_n}{n} \in I_n\right) \simeq 0,95$ à $0,01$ près (ce qui est vrai si $n \geqslant 200$).

Vérifier que si $I_n \subset J_n$ alors $P\left(\frac{X_n}{n} \in J_n\right) \geqslant 0,94$.

(5) Traduire sous forme de deux inégalités l'inclusion précédente.

(6) Application numérique :

   (a) On suppose que $p = 0,5$ et $n = 1000$. Quelle est la taille minimale des deux RU ?

   (b) On suppose que $p = 0,3$ et $n = 1000$. Quelle est la taille minimale des deux RU ?

   (c) On suppose que $p = 0,5$, $N_1 = N_2 = 500$. En utilisant une inéquation du second degré, déterminer le nombre maximum d'étudiants qui pourront trouver une place.

(1) X_n *suit une loi binomiale de paramètres* n *et* p.

(2) *Question de cours :* $I_n = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.

(3) $J_n = \left[1 - \frac{N_2}{n}; \frac{N_1}{n}\right]$. *Il faut avoir* $P\left(\frac{X_n}{n} \in J_n\right) \geqslant 0,95$ *à* $0,01$ *près soit* $P\left(\frac{X_n}{n} \in J_n\right) \geqslant 0,94$

(4) *Clair.*

(5) $I_n \subset J_n \Leftrightarrow \begin{cases} 1 - \frac{N_2}{n} \leqslant p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{N_1}{n} \end{cases} \Leftrightarrow \begin{cases} n(1-p) + 1,96\sqrt{n}\sqrt{p(1-p)} \leqslant N_2 \\ np + 1,96\sqrt{n}\sqrt{p(1-p)} \leqslant N_1 \end{cases}$.

(6) (a) *On a la même inéquation donnant* $N_1 \geqslant 531$ *et* $N_2 \geqslant 531$.

   (b) *On obtient* $N_1 \geqslant 329$ *et* $N_2 \geqslant 729$.

   (c) *On résout* $0,5n + 0,98\sqrt{n} - 500 \leqslant 0$ *ce qui donne* $n \leqslant 938$.