Chapitre 16

Échantillonnage

I Exercices

16.1 Loi binomiale, loi normale et échantillonnage

Exercice 16.1

Une société de démarchage téléphonique sait que, en contactant une personne, la probabilité de réussir à vendre un produit donné est 0,4.

1. Un commercial contacte 30 clients successivement. On appelle X la variable aléatoire égale au nombre de succès.
   a) Quelle est la loi de cette variable aléatoire ?
   b) Calculer μ l'espérance de cette variable aléatoire, et son écart-type σ arrondi au centième près.
   c) Calculer $p(\mu - 1,96\sigma \leqslant X \leqslant \mu + 1,96\sigma)$
   Indications :
   - calculer d'abord $\mu - 1,96\sigma$ et $\mu + 1,96\sigma$;
   - déterminer les valeurs de X qui sont entre ces deux nombres ;
   - calculer la probabilité demandée, en utilisant `binomFRép`, en sachant que : $p(0 \leqslant X \leqslant k) =$ `binomFRép(n,p,k)`.
2. Y est la variable aléatoire qui suit la loi normale d'espérance μ et d'écart-type σ.
   Calculer $p(\mu - 1,96\sigma \leqslant Y \leqslant \mu + 1,96\sigma)$.
3. **Généralisation**
   Un commercial contacte n clients, avec à chaque client une probabilité de succès p, et X est le nombre de succès. L'ensemble des n clients est en fait un échantillon dans l'ensemble de tous les clients.
   On admet que : $p(\mu - 1,96\sigma \leqslant X \leqslant \mu + 1,96\sigma)$ est proche de 0,95
   Faisons maintenant le lien avec l'intervalle de fluctuation.
   Puisque X est le nombre de succès pour n répétitions, la fréquence de succès est : $f = \dfrac{X}{n}$
   Démontrer que :
   $$\mu - 1,96\sigma \leqslant X \leqslant \mu + 1,96\sigma \iff p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant f \leqslant p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

16.2 Intervalle de fluctuation

Exercice 16.2

Environ 51 % de la population française ayant un emploi fait partie de la catégorie socio-professionnelle « employés et ouvriers ».

Au concours de l'ENA en 2009, 17 admis sur 81 ont des parents appartenant à cette catégorie.

1. Dans un échantillon de 81 personnes choisies au hasard dans la population française ayant un emploi, on peut déterminer l'intervalle de fluctuation asymptotique au seuil de confiance de 95 % de la fréquence de personnes faisant partie de cette catégorie.
   a) Vérifier les conditions d'applications d'application de l'intervalle de fluctuation.
   b) Calculer cet intervalle.
2. Peut-on considérer qu'en 2009, les admis dont les parents font partie de la catégorie « employés et ouvriers », sont sous représentés à l'ENA ?

Exercice 16.3

Une ville offre pour Noël un panier cadeaux aux seniors (personnes de plus de 60 ans). Dans cette ville, 17 % des habitants ont plus de 60 ans.

On distribue ce panier dans un quartier de 550 habitants.

1. Déterminer l'intervalle de fluctuation asymptotique au seuil de confiance de 95 % de la fréquence de seniors dans ce quartier.
2. Combien de paniers faut-il prévoir ?

Exercice 16.4

La documentation technique d'une machine à fabriquer des pilules indique que la proportion de pilules défectueuses fabriquées par cette machine est de 1 pour 1 000.

Sur un échantillon de 10 000 pilules, 15 pilules sont défectueuses.

La proportion annoncée semble-t-elle respectée ? On pourra utiliser un intervalle de fluctuation au seuil de 95 %.

16.3 Intervalle de confiance

Exercice 16.5

1. Un sondage effectué auprès de 1 000 personnes annonce qu'un candidat à une élection a 51,24 % d'intentions de vote.
   a) Calculer l'intervalle de confiance au seuil de 95 % de la proportion de voix favorables à ce candidats.
   b) Peut-on affirmer, avec un niveau de confiance de 95 % que ce candidat sera élu ?
2. Mêmes questions **a)** et **b)** pour une autre élection, où un sondage effectué auprès de 1 000 personnes annonce qu'un autre candidat à une élection a 58,62 % d'intentions de vote.
3. **Précision des sondages**
   a) Quand un sondage est effectué auprès de 1 000 personnes, quelle est la précision en pourcentage de ce sondage, autrement dit le résultat est valable à plus ou moins quel pourcentage ?
   b) Que pensez-vous de la précision en pourcentage à 0,01 % près des deux sondages précédents ?

c) Un sondage « sortie des urnes » est effectué le jour de l'élection auprès de 100 000 votants entre 18 h et 20 h pour donner le résultat dans les médias à 20 h. Quelle est alors la précision d'un tel sondage ?

Exercice 16.6

Dans la production d'une usine automobiles, 120 voitures sont vérifiées et 94 n'ont pas de défaut.

1. Dans quel intervalle est la proportion p de voitures sans défaut dans la production de cette usine ?
2. Est-on certain que la proportion p appartienne à cet intervalle ? Justifier.

Exercice 16.7

Un nouveau traitement est testé dans un hôpital sur 200 patients et échoue 37 fois.

1. Donner un intervalle permettant d'estimer le pourcentage de réussite de ce traitement.
2. Sur quel effectif de patients faudrait-il tester ce traitement pour estimer le pourcentage de réussité à 1 % près ?

16.4 Exercices de bac

Exercice 16.8 (Baccalauréat S, Liban, mai 2015, ex. 4)

En prévision d'une élection entre deux candidats A et B, un institut de sondage recueille les intentions de vote de futurs électeurs.

Parmi les 1 200 personnes qui ont répondu au sondage, 47 % affirment vouloir voter pour le candidat A et les autres pour le candidat B.

Compte-tenu du profil des candidats, l'institut de sondage estime que 10 % des personnes déclarant vouloir voter pour le candidat A ne disent pas la vérité et votent en réalité pour le candidat B, tandis que 20 % des personnes déclarant vouloir voter pour le candidat B ne disent pas la vérité et votent en réalité pour le candidat A.

On choisit au hasard une personne ayant répondu au sondage et on note :

- A l'évènement « La personne interrogée affirme vouloir voter pour le candidat A » ;
- B l'évènement « La personne interrogée affirme vouloir voter pour le candidat B » ;
- V l'évènement « La personne interrogée dit la vérité ».

1. Construire un arbre de probabilités traduisant la situation.
2. a) Calculer la probabilité que la personne interrogée dise la vérité.
   b) Sachant que la personne interrogée dit la vérité, calculer la probabilité qu'elle affirme vouloir voter pour le candidat A.
3. Démontrer que la probabilité que la personne choisie vote effectivement pour le candidat A est $0,529$.
4. L'institut de sondage publie alors les résultats suivants :

52,9 % des électeurs* voteraient pour le candidat A. *estimation après redressement, fondée sur un sondage d'un échantillon représentatif de 1 200 personnes.

   Au seuil de confiance de 95 %, le candidat A peut- il croire en sa victoire ?
5. Pour effectuer ce sondage, l'institut a réalisé une enquête téléphonique à raison de 10 communications par demi-heure. La probabilité qu'une personne contactée accepte de répondre à cette enquête est $0,4$.

   L'institut de sondage souhaite obtenir un échantillon de 1 200 réponses.

   Quel temps moyen, exprimé en heures, l'institut doit-il prévoir pour parvenir à cet objectif ?

Exercice 16.9 (Baccalauréat S, Métropole-Réunion, juin 2015, ex. 1)
Les résultats des probabilités seront arrondis à 10^{-3} *près.*

Partie 1

1. Soit X une variable aléatoire qui suit la loi exponentielle de paramètre λ, où λ est un réel strictement positif donné.
On rappelle que la densité de probabilité de cette loi est la fonction f définie sur $[0\ ;\ +\infty[$ par

$$f(x) = \lambda \mathrm{e}^{-\lambda x}.$$

a) Soit c et d deux réels tels que $0 \leqslant c < d$.
Démontrer que la probabilité $P(c \leqslant X \leqslant d)$ vérifie $P(c \leqslant X \leqslant d) = \mathrm{e}^{-\lambda c} - \mathrm{e}^{-\lambda d}$.

b) Déterminer une valeur de λ à 10^{-3} près de telle sorte que la probabilité $P(X > 20)$ soit égale à 0,05.

c) Donner l'espérance de la variable aléatoire X

Dans la suite de l'exercice on prend $\boldsymbol{\lambda = 0,15}$.

d) Calculer $P(10 \leqslant X \leqslant 20)$.

e) Calculer la probabilité de l'évènement $(X > 18)$.

2. Soit Y une variable aléatoire qui suit la loi normale d'espérance 16 et d'écart type $1,95$.

a) Calculer la probabilité de l'évènement $(20 \leqslant Y \leqslant 21)$.

b) Calculer la probabilité de l'évènement $(Y < 11) \cup (Y > 21)$.

Partie 2

Une chaîne de magasins souhaite fidéliser ses clients en offrant des bons d'achat à ses clients privilégiés. Chacun d'eux reçoit un bon d'achat de couleur verte ou rouge sur lequel est inscrit un montant.

Les bons d'achats sont distribués de façon à avoir, dans chaque magasin, un quart de bons rouges et trois quarts de bons verts.

Les bons d'achat verts prennent la valeur de 30 euros avec une probabilité égale à $0,067$ ou des valeurs comprises entre 0 et 15 euros avec des probabilités non précisées ici.

De façon analogue, les bons d'achat rouges prennent les valeurs 30 ou 100 euros avec des probabilités respectivement égales à $0,015$ et $0,010$ ou des valeurs comprises entre 10 et 20 euros avec des probabilités non précisées ici.

1. Calculer la probabilité d'avoir un bon d'achat d'une valeur supérieure ou égale à 30 euros sachant qu'il est rouge.

2. Montrer qu'une valeur approchée à 10^{-3} près de la probabilité d'avoir un bon d'achat d'une valeur supérieure ou égale à 30 euros vaut $0,057$.

Pour la question suivante, on utilise cette valeur.

3. Dans un des magasins de cette chaîne, sur 200 clients privilégiés, 6 ont reçu un bon d'achat d'une valeur supérieure ou égale à 30 €.

Le directeur du magasin considéré estime que ce nombre est insuffisant et doute de la répartition au hasard des bons d'achats dans les différents magasins de la chaîne.
Ses doutes sont-ils justifiés ?

II Cours

16.1 Rappels sur l'échantillonnage

On s'intéresse à un caractère dans une population. La proportion de ce caractère dans cette population est noté p. Un échantillon est constitué d'individus de cette population, en général choisis au hasard. Son effectif est noté n et il est en général beaucoup plus réduit que l'effectif de la population. La fréquence du caractère dans l'échantillon est notée f.

16.2 Intervalle de fluctuation

Le programme de mathématiques de terminale S indique qu'un élève doit connaître l'intervalle de fluctuaton asymptotique au seuil de 95 % (voir propriété ci-dessous).

Situation : la proportion p du caractère dans la population est donnée.

Plus précisément la propriété énoncée ci-dessous sera utilisée lorsque la proportion p du caractère dans la population est donnée, soit parce qu'elle est connue de manière certaine, soit parce que cette proportion p est annoncée, et qu'on se demande si la valeur de p est plausible ou non. L'effectif n de l'échantillon et la fréquence f du caractère dans l'échantillon sont connus (il faut parfois calculer f). On calcule alors l'intervalle de fluctuation pour tirer des conclusions, soit sur l'échantillon, soit sur la valeur de p.

À l'aide de la loi binomiale et de la loi normale, on démontre la propriété ci-dessous.

Propriété 16.1

p est la proportion d'un caractère étudié dans une population, n est l'effectif d'un échantillon de cette population, et f est la fréquence du caractère dans cet échantillon.
Si $n \geqslant 30$, $np \geqslant 5$, $n(1-p) \geqslant 5$,
alors, il y a une probabilité proche de 0,95 que la fréquence f appartienne à l'intervalle
$$\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$
Cet intervalle est appelé **intervalle de fluctuaton asymptotique au seuil de 95 %**.

16.3 Estimation – Intervalle de confiance

Le programme de mathématiques de terminale S indique qu'un élève doit savoir

- *estimer une proportion inconnue à partir d'un échantillon ;*
- *déterminer une taille d'échantillon suffisante pour obtenir, avec une précision donnée, une estimation d'une proportion au niveau de confiance 0,95.*

Situation : la proportion p du caractère dans la population n'est pas donnée.

La propriété énoncée ci-dessous sera utilisée lorsque la proportion p du caractère dans la population est inconnue. La fréquence f du caractère dans l'échantillon est donnée ou il faut la calculer. On calcule alors l'intervalle de confiance et cela sert à **estimer** la valeur de p.

Propriété 16.2

p est la proportion d'un caractère étudié dans une population, n est l'effectif d'un échantillon de cette population, et f est la fréquence du caractère dans cet échantillon.
Si $n \geqslant 30$, $np \geqslant 5$, $n(1-p) \geqslant 5$,
alors, il y a une probabilité supérieure ou égale à 0,95 que la proportion p appartienne à l'intervalle $\left[f - \dfrac{1}{\sqrt{n}}, f + \dfrac{1}{\sqrt{n}}\right]$.
Cet intervalle est appelé **intervalle de confiance au niveau de confiance 0,95**.