STT1000

Chapitre 4 Estimation ponctuelle

Exercices

Estimateurs sans biais

4.1 Soit $X_1, \ldots, X_n$ un échantillon aléatoire tiré d'une population de moyenne μ et de variance σ^2. Considérons l'ensemble des estimateurs de la forme $\hat{\mu} = \sum_{i=1}^{n} a_i X_i$ où $a_1, \ldots, a_n$ sont des constantes. Quelles conditions les a_i devraient-elles satisfaire pour que $\hat{\mu}$ soit sans biais ?

4.2 Soit $X_1, \ldots, X_n$ un échantillon aléatoire tiré d'une population de densité

$$f(x) = \frac{x}{\theta^2} e^{-x/\theta}; \quad x > 0, \theta > 0.$$

On cherche un estimateur sans biais de θ et on considère $\bar{X}$ comme un premier essai. Montrer que $\bar{X}$ est biaisé, et montrez comment modifier cet estimateur pour en obtenir un qui ne l'est pas.

4.3 Soit $X_1, \ldots, X_n$ un échantillon aléatoire tiré d'une population de densité

$$f(x) = \frac{1}{\Gamma(\theta)2^{\theta}} x^{\theta-1} e^{-x/2}; \quad x > 0, \theta > 0.$$

On cherche un estimateur sans biais de θ et on considère $\bar{X}$ comme un premier essai. Montrer que $\bar{X}$ est biaisé, et montrer comment modifier cet estimateur pour en obtenir un qui ne l'est pas.

4.4 Deux recherches indépendantes font état d'échantillonnages effectués auprès d'une même population. Les seules données présentées sont les moyennes $\bar{X}_1$ et $\bar{X}_2$ et les tailles des échantillons n_1 et n_2.

Déterminer la valeur k telle que $k(\bar{X}_1 - \bar{X}_2)^2$ est un estimateur sans biais de la variance σ^2 de la population.

4.5 Soit $X_1, \ldots, X_n$ un échantillon aléatoire d'une population de loi de $\mathcal{G}(p)$.

a) Montrer que $\bar{X}$ est un estimateur sans biais de $1/p$.

b) Montrer que $\dfrac{1 - 1/n}{\bar{X} - 1/n}$ est sans biais pour p. [*Suggestion*: Se servir du fait que la somme des probabilités sur toutes les valeurs d'une variable aléatoire (de loi binomiale négative dans le cas présent) est égale à 1].

Variance d'un estimateur

4.6 Soit $X_1, \ldots, X_n$ un échantillon aléatoire tiré d'une population de moyenne μ et de variance σ^2. Considérons l'ensemble des estimateurs de la forme $\hat{\mu} = \sum_{i=1}^{n} a_i X_i$ où $a_1, \ldots, a_n$ sont des constantes. Montrer que celui qui a la plus petite variance est $\bar{X}$. [*Suggestion*: Déduisez de l'inégalité $\sum_{i=1}^{N} \left(a_i - \frac{1}{n}\right)^2 \geq 0$ que $\sum a_i^2 \geq 1/n$].

4.7 Soit $X_1, \ldots, X_n$ un échantillon d'une population $\mathcal{N}(\mu ; \sigma^2)$. Dans cet exercice, on suppose μ connu, et on compare l'estimateur S^2 à l'estimateur $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$.

a) Montrez que $n\hat{\sigma}^2/\sigma^2 = \displaystyle\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$ donc que $\mathrm{E}\left(\dfrac{n\hat{\sigma}^2}{\sigma^2}\right) = n$ et par conséquent $\hat{\sigma}^2$ est un estimateur sans biais.

b) Montrer que $\mathrm{Var}\left(\dfrac{n\hat{\sigma}^2}{\sigma^2}\right) = 2n$, et par conséquent $\mathrm{Var}(\hat{\sigma}^2) = \dfrac{2\sigma^4}{n}$.

c) Utiliser le fait que $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ pour montrer que $\mathrm{Var}[S^2] = \dfrac{2\sigma^4}{n-1}$.

d) Quel est le meilleur estimateur, $\hat{\sigma}^2$ ou S^2 ?

4.8 Soit $X_1, \ldots, X_n$ un échantillon aléatoire simple d'une population de moyenne μ. Déterminer les variances des trois estimateurs $T_1 = X_1$, $T_2 = 2X_1 - X_2$ et $T_3 = 2\sum_{i=1}^{n} \dfrac{i}{n(n+1)} X_i$. En quel sens $\bar{X}$ est-il préférable aux deux autres?

4.9 Soit $\bar{X}_1, \ldots, \bar{X}_k$ les moyennes de k échantillons indépendants de tailles $n_1, \ldots, n_k$, tirés tous de la même population de moyenne μ et de variance σ^2.

a) Pour estimer μ on considère l'estimateur linéaire $\hat{\mu} = a_1\bar{X}_1 + a_2\bar{X}_2 + \ldots + a_k\bar{X}_k$, où $a_1, \ldots, a_k$ sont des constantes. Quelle(s) condition(s) les a_i doivent-elles satisfaire pour que $\hat{\mu}$ soit sans biais ?

b) Parmi tous les estimateurs linéaires sans biais, montrer que celui qui a la plus petite variance est celui où $a_i = n_i/n$ où $n = \sum_{i=1}^{n} n_i$. [*Indice*: montrer que $\sum_{i=1}^{n} \dfrac{a_i^2}{n_i} \geq \dfrac{1}{n}$ à partir de l'énoncé $\sum_{i=1}^{n} n_i\left(\dfrac{a_i}{n_i} - \dfrac{1}{n}\right)^2 \geq 0$.]

4.10 Une population est constituée de deux sous-populations, de moyennes μ_1 et μ_2 et de variances connues σ_1^2 et σ_2^2. On doit prélever un échantillon de chacune des deux sous-populations. Le budget impose la contrainte $n = n_1 + n_2 = 100$ sur les tailles n_1 et n_2 des deux échantillons.

a) Montrer que si l'objectif est d'estimer $\mu_1-\mu_2$ par la différence $\bar{X}_1 - \bar{X}_2$ des deux moyennes échantillonnales, le meilleur choix de n_1 et n_2 est $n_1 = \dfrac{100\sigma_1}{\sigma_1+\sigma_2}$ et $n_2 = \dfrac{100\sigma_2}{\sigma_1+\sigma_2}$?

b) Les tailles relatives des deux sous-populations sont de 20 % et 80 %, de sorte que la moyenne de la population entière est $\mu = 0{,}2\mu_1 + 0{,}8\mu_2$. Montrer que si l'objectif est d'estimer μ (par $0{,}2\bar{X}_1 + 0{,}8\bar{X}_2$), le meilleur choix de n_1 et n_2 est $n_1 = \dfrac{0{,}2\sigma_1}{0{,}2\sigma_1 + 0{,}8\sigma_2}$ et $n_2 = \dfrac{0{,}8\sigma_2}{0{,}2\sigma_1 + 0{,}8\sigma_2}$?

4.11 Si $X \sim \mathcal{B}(n\,;p)$ et $\hat{p} = \dfrac{X}{n}$, montrez que $\mathrm{Var}(\hat{p}) \leq \dfrac{1}{4n}$.

4.12 Le nombre de fautes de frappe dans une page de livre est une variable de loi de Poisson de paramètre γ. Deux réviseurs vérifient l'un 10 pages, l'autre 20 pages. Soit X le nombre total de fautes détectées par le premier (sur 10 pages); et Y le nombre total de fautes détectées par le deuxième (20 pages).

a) Déterminer la condition que doivent satisfaire a et b pour que l'estimateur $\hat{\gamma} = aX + bY$ soit sans biais pour γ [c'est-à-dire, exprimer a en fonction de b].

b) De tous les estimateurs sans biais de la forme $\hat{\gamma} = aX + bY$, quel est le meilleur? Évaluez $\hat{\gamma}$ si $X = 3$ et $Y = 5$.

4.13 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi uniforme sur $[0\,;\beta]$. Considérez les deux estimateurs suivants : $\hat{\beta}_1 = 2\bar{X}$ et $\hat{\beta}_2 = \max\{X_1\,; \ldots; X_n\}$.

a) Montrer que $\hat{\beta}_1$ est sans biais.

b) Montrer que $\hat{\beta}_2$ est biaisé, déterminez-en le biais, et déterminer un facteur c tel que $\hat{\beta}_3 = c\hat{\beta}_2$ soit sans biais.

c) Déterminer la variance de $\hat{\beta}_1$ et la variance de $\hat{\beta}_3$ et dites lequel des deux estimateurs est meilleur.

Méthode des moments

4.14 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi gamma de paramètres α et β.

a) Montrer que la méthode des moments donne les estimateurs suivants :

$$\hat{\alpha} = \frac{n\bar{X}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \quad \text{et} \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\bar{X}}.$$

b) On choisit 12 jours au hasard et on note la consommation d'énergie (en millions de kilowattheures) ces jours-là :
`6,4; 4,5; 10,8; 7,2; 6,8; 4,9; 3,5; 16,3; 4,8; 7,0; 8,8; 5,5.`

Estimez α et β par la méthode des moments.

4.15 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur [0 ; β]. Déterminer par la méthode des moments l'estimateur de β.

4.16 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur [α ; 1]. Déterminer par la méthode des moments l'estimateur de α.

4.17 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de densité

$$f(x) = \begin{cases} \frac{2}{\beta}(1 - x/\beta) & 0 < x < \beta \\ 0 & \text{ailleurs} \end{cases}.$$

Déterminer par la méthode des moments l'estimateur de β.

4.18 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi gamma de paramètres α = 2 (connu) et β inconnu. Déterminer par la méthode des moments l'estimateur de β.

4.19 Soit X_1; …;X_n un échantillon aléatoire simple d'une population de loi géométrique de paramètre p. Déterminer par la méthode des moments l'estimateur de p.

4.20 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de densité $f(x) = \begin{cases} \frac{1}{\beta}e^{-(x-\alpha)/\beta} & x > \alpha \\ 0 & \text{ailleurs} \end{cases}$

Déterminer par la méthode des moments de α et de β.

4.21 Un inspecteur doit estimer la proportion p de ménages munis d'un détecteur de fumée. Il décide donc d'inspecter des ménages et de continuer jusqu'au moment où il obtient un **10^e^** ménage avec détecteur de fumée. Sachant que ce 10^e^ ménage est le **50^e^** inspecté, déterminer l'estimateur de p par la méthode des moments et estimer p.

4.22 Supposons que le nombre de taches dans une plaque d'émail est une variable de loi de Poisson de paramètre λ. Dans un échantillon de n plaques vous comptez le nombre de taches dans chacune des plaques.

a) Supposons que vous ne retenez que l'information suivante : Le nombre X de plaques n'ayant *aucune* tache. Déterminer l'estimateur par les moments de λ.

b) Dans un échantillon de 25 plaques, on en trouve 18 sans aucune tache. Estimer λ.

c) Supposons maintenant qu'en fait vous avez retenu toute l'information disponible, soit les valeurs X_1, X_2, … , X_n, X_i étant le nombre taches dans la plaque i. Déterminer l'estimateur par les moments de λ.

d) Estimer λ à l'aide de l'estimateur en b) sachant que 25 plaques ont été échantillonnées avec les résultats suivants : 18 des plaques n'avaient aucune tache, 5 en avaient une, et 2 en avaient 4.

Maximum de vraisemblance

4.23 Une observation $X \sim \mathcal{B}(n\,;p)$ est en fait un résumé des observations élémentaires: X est la somme de n variables aléatoires X_1, X_2, …, X_n, chacune prenant l'une des valeurs 0 ou 1. Donc X seule ne donne pas d'information sur l'ordre dans lequel les succès se sont produits. Est-ce que cette information perdue aurait pu être utile ?

a) Vérifier que la fonction de vraisemblance de ces observations est $L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$ et que l'estimateur du maximum de vraisemblance est encore X/n.

b) Montrer que la distribution conditionnelle de $[X_1\,;X_2\,;\dots;X_n]$ étant donné $X = x$ est $p(x_1\,;x_2\,;\dots;x_n) = \dfrac{1}{\binom{n}{x}}$ pour tout vecteur $[x_1\,;x_2\,;\dots;x_n]$ dont la somme est x. On remarque que cette distribution conditionnelle ne dépend pas de p et donc ne contient aucune information sur ce paramètre.

4.24 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de densité

$$f(x) = \begin{cases} \dfrac{2}{\beta}(1-x/\beta) & 0 < x < \beta \\ 0 & \text{ailleurs} \end{cases}.$$

Déterminer l'estimateur de β par la méthode du maximum de vraisemblance [Il suffira d'indiquer quelle équation doit être résolue. La solution n'est pas facile à trouver].

4.25 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi géométrique de paramètre p. Déterminer par la méthode du maximum de vraisemblance l'estimateur de p.

4.26 On tire deux échantillons aléatoires simples, de tailles n_1 et n_2, d'une même population normale de moyenne μ et de variance σ^2. Les seules informations retenues sont les moyennes $\overline{X}_1$ et $\overline{X}_2$. Montrez que les estimateurs $\hat{\mu}$ et $\hat{\sigma}^2$ du maximum de vraisemblance μ et σ^2, respectivement, sont $\hat{\mu} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n}$, où $n = n_1+n_2$; et $\hat{\sigma}^2 = \dfrac{n_1(\overline{X}_1-\hat{\mu})^2 + n_2(\overline{X}_2-\hat{\mu})^2}{2}$.

4.27 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur $[0\,;\beta]$. Déterminer par la méthode du maximum de vraisemblance l'estimateur de β. [rép. $\hat{\beta} = \max\{X_1; \dots; X_n\}$]

4.28 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur $[\alpha\,;1]$. Déterminer par la méthode du maximum de vraisemblance l'estimateur de α.

4.29 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de loi uniforme sur $[\alpha\,;\beta]$. Déterminer par la méthode du maximum de vraisemblance les estimateurs de α et de β .

4.30 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de densité $f(x) = \begin{cases} \dfrac{1}{\beta}e^{-(x-\alpha)/\beta} & x > \alpha \\ 0 & \text{ailleurs} \end{cases}$. Déterminer par la méthode du maximum de vraisemblance les estimateurs de α et de β.

4.31 Soit X_1; …; X_n un échantillon aléatoire simple d'une population de fonction de densité $f(x) = 2x/\theta^2$ pour $0 < x < \theta$ et zéro ailleurs.

a) Déterminer un estimateur sans biais de la forme $\hat{\theta}_1 = c\bar{X}$.

b) Déterminer la variance de $\hat{\theta}_1$

c) Montrer que l'estimateur par la méthode du maximum de vraisemblance est $X_{(n)}$

d) Déterminer la fonction de densité de $Y = X_{(n)}$, son espérance et sa variance

e) Modifier $X_{(n)}$ de façon à obtenir un estimateur sans biais $\hat{\theta}_2$.

f) Déterminer la variance de $\hat{\theta}_2$

g) Montrer que $\hat{\theta}_2$ est le meilleur des deux estimateurs.

f) Montrer que $\dfrac{X_{(n)}^2}{4n^2}$ est un estimateur sans biais de $\mathrm{Var}(\hat{\theta}_2)$

g) Montrer que la variance de $\hat{\theta}_1$ peut être estimée sans biais par $\dfrac{n+1}{8n^2}X_{(n)}^2$ ainsi que par $\dfrac{9}{4n}S^2$

f) On tire un échantillon de taille 4 de cette population. Les données de l'échantillon sont
1 ; 3 ; 3 ; 4 ; 5 ; 6 ; 8 ; 12.

Estimer θ par les deux méthodes.

4.32 On tire deux échantillons aléatoires simples, de tailles n_1 et n_2, d'une même population normale de moyenne μ et de variance σ^2. Les seules informations retenues sont les moyennes $\bar{X}_1$ et $\bar{X}_2$. Montrez que les estimateurs $\hat{\mu}$ et $\hat{\sigma}^2$ du maximum de vraisemblance μ et σ^2, respectivement, sont $\hat{\mu} = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n}$, où $n = n_1+n_2$; et $\hat{\sigma}^2 = \dfrac{n_1(\bar{X}_1 - \hat{\mu})^2 + n_2(\bar{X}_2 - \hat{\mu})^2}{2}$.

4.33 On tire un échantillon de taille 4 d'une population de densité $f(x) = (1/\beta^2)xe^{-x/\beta}$, pour $x > 0$ et zéro ailleurs. Les données d'un échantillon de taille 4 sont $X_1 = 3$; $X_2 = 4$; $X_3 = 6$; $X_4 = 12$. Estimer β par la méthode du maximum de vraisemblance

4.34 La teneur X en mercure des poissons d'un étang très pollué est une variable de densité

$$f(x) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{ailleurs} \end{cases}, \qquad \theta > 0\,.$$

Soit $X_1, X_2, \ldots, X_n$ les teneurs en mercure de n poissons tirés dans l'étang.

a) Déterminer l'estimateur $\hat{\theta}$ de θ par la méthode du maximum de vraisemblance

b) Déterminer l'estimateur par la moyenne. On verra qu'elle ne coïncide pas avec $\hat{\theta}$.

c) Transformer les données par $Y_i = -\ln X_i$, et déterminer l'estimateur de θ par les deux méthodes. Est-ce qu'elles coïncident cette fois-ci?

4.35 Cinq pièces identiques font partie d'un appareil électronique. La durée de vie (en nombre d'heures) de chacune est une variable de loi exponentielle de moyenne β, et les 5 durées de vie sont indépendantes. L'appareil tombe en panne dès que l'une des 5 pièces cesse de fonctionner. Vous observez que l'appareil cesse de fonctionner après exactement 8 heures. Utiliser cette seule information pour déterminer l'estimateur du maximum de vraisemblance de β. Évaluer l'estimateur numériquement.

4.36 Un chercheur en médecine a besoin d'un sujet possédant un gène particulier, très rare. Afin d'en trouver un, il tire au hasard des individus dans une certaine (très grande) population jusqu'au moment où il tombe sur une personne possédant ce gène. Il lui a fallu tirer **200** personnes dans la population pour obtenir ce sujet. Déterminer l'estimateur du maximum de vraisemblance $\hat{p}$ de p, la proportion d'individus dans la population qui possèdent le gène en question. L'estimateur est-il sans biais ?

4.37 Considérons une population de loi de Pareto, $f(x) = \alpha x^{-\alpha-1}$, $x > 1$.

a) Déterminer l'estimateur du maximum de vraisemblance de α étant donné un échantillon de taille n d'une population de densité f. L'estimateur est-il sans biais ?

b) Supposons que les salaires (en dollars) d'un échantillon de professeurs du secondaires sont les suivants :

31 332; 23 038; 10 395; 29 432; 59 329; 45 139; 13 581; 15 016 ;

10 582; 17 295; 22 154; 23 280; 14 013; 20 799; 17 000.

Supposez que les salaires, exprimés en dizaines de milliers de dollars, sont de loi de Pareto de paramètre α. Estimer α.

4.38 Une observation X de loi binomiale négative de paramètres r et p est en fait un résumé des observations élémentaires: X est la somme de r variables aléatoires X_1, X_2, …, X_n, chacune de loi géométrique de paramètre p. Donc X seule ne donne pas d'information sur les valeurs individuelles de X_i. Est-ce que cette information perdue aurait pu être utile ?

a) Vérifier que la fonction de vraisemblance de ces observations est $L(p) = (1-p)^{x-r}\,p^r$

$\prod_{i=1}^{r}[(1-p)^{x_i-1}p] = (1-p)^{\sum_{i=1}^{n}(x_i-1)}p^r = (1-p)^{x-r}\,p^r$ et que l'estimateur du maximum de vraisemblance est encore r/X.

b) Montrer que la distribution conditionnelle de $[X_1 ; X_2 ; … ; X_n]$ étant donné $X = x$ est $p(x_1 ; x_2 ; … ; x_n) = \dfrac{1}{\binom{x-1}{r-1}}$ pour tout vecteur $[x_1 ; x_2 ; … ; x_n]$ dont la somme est x. On remarque que cette distribution conditionnelle ne dépend pas de p et donc ne contient aucune information sur ce paramètre.

Problèmes divers

4.39 Le nombre d'arrivées à une salle d'urgence dans un intervalle de temps $(0 ; t)$ les mercredis matin est une variable de loi de Poisson de paramètre $\lambda = \gamma t$.

a) Afin d'estimer γ (le taux d'arrivées par unité de temps), on compte X, le nombre d'arrivées dans un intervalle de longueur t. Soit $\hat{\gamma}_1 = \dfrac{X}{t}$ un estimateur de γ. Montrer que $\hat{\gamma}_1$ est sans biais et que $\mathrm{Var}(\hat{\gamma}_1) = \dfrac{\gamma}{t}$.

b) Supposons qu'on procède différemment : on observe les arrivées et on note Y, le temps écoulé au moment de la r^e arrivée.

i) Montrer que $\hat{\gamma}_2 = \dfrac{r-1}{Y}$ est sans biais.

ii) Vérifier que si $r > 2$, la variance de $\hat{\gamma}_2$ est $\mathrm{Var}(\hat{\gamma}_2) = \dfrac{\gamma^2}{r-2}$.

iii) Montrer que $\mathrm{Var}(\hat{\gamma}_2) \le \mathrm{Var}(\hat{\gamma}_1)$ si et seulement si $\gamma t \le r - 2$.

4.40 *Estimation du paramètre d'une loi hypergéométrique* Un intervalle de confiance peut être défini comme l'ensemble de toutes les valeurs crédibles du paramètre au vu de l'échantillon. L'exemple ci-dessous illustre la façon de déterminer si une valeur du paramètre N_1 d'une loi hypergéométrique est crédible ou pas, donc à inclure ou pas dans l'intervalle de confiance.

L'acheteur potentiel d'un immeuble de N = 100 appartements tire un échantillon de n = 30 appartements.

Selon le vendeur, 35 appartements ont des cuisines rénovées. Mais dans l'échantillon, 5 seulement des 30 appartements le sont.

a) Soit N_1 le nombre d'appartements avec cuisine rénovée dans l'immeuble. Si l'affirmation du vendeur (que N_1 = 35) est vraie, quelle est l'espérance du nombre de cuisines rénovées dans l'échantillon ? 10,5

b) Calculer $P(X \leq 5)$ sous la supposition que que $N_1 = 35$. L'affirmation du vendeur est-elle crédible (à la lumière du résultat) ?

c) Quelles valeurs de N_1 sont crédibles ?

d) Mais est-ce que toutes les valeurs de N_1 inférieures à 30 sont crédibles ? Certaines seraient-elles trop petites ?

4.41 Cet exemple illustre la façon de déterminer si une valeur du paramètre p d'une loi binomiale est crédible ou pas. Vous tirez un échantillon de 20 pièces fabriquées d'un processus dont le taux de défectuosités (le pourcentage de pièces défectueuses) est censé être de 2 %. Vous observez 4 pièces défectueuses.

a) Sous l'hypothèse que le taux de défectuosités est bien de 2 %, quelle est l'espérance du nombre de pièces défectueuses dans votre échantillon ?

b) Sous l'hypothèse que le taux de défectuosités est bien de 2 %, quelle est la probabilité d'avoir 4 pièces ou plus dans votre échantillon ?

c) À la lumière de la probabilité calculée en b), que concluez-vous à propos de l'hypothèse que p = 2 % ?

d) Est-ce que l'hypothèse $p = 0{,}05$ vous semble plus vraisemblable ?

e) Lesquelles des valeurs suivantes de p sont vraisemblables, compte tenu de l'observation $X = 4$: 0,05; 0,06; 0,07; 0,08; 0,09; 0,10 ?

f) On décide de tolérer un taux de défectuosité de 10 % et on décide aussi de rejeter le lot si le nombre de pièces défectueuses dans l'échantillon est supérieur ou égal à 4. Si en fait le lot est juste acceptable ($p = 0{,}10$). Quelle est la probabilité de le rejeter ?

g) On change de critère car on trouve la probabilité calculée en i) trop élevée : on rejettera le lot seulement si $X \geq 5$. Quelle est la probabilité de rejeter un lot dans lequel 10 % des pièces sont défectueuses ?

h) La probabilité calculée en ii) pourrait être considérée suffisamment faible ou pas, ce sont les personnes impliquées qui en décident. On aura normalement fixé un seuil α (la probabilité maximale de rejet d'un lot acceptable.) Supposons qu'on pose $\alpha = 0{,}01$. À partir de quelle valeur de X devrait-on rejeter le lot ?

4.42 Supposons que le nombre X de rongeurs qui passent dans votre cours en une nuit est une variable de loi de Poisson de paramètre λ.

a) Si vous observez $X = 2$, la valeur $\lambda = 5$ est-elle vraisemblable ?

b) Parmi les valeurs suivantes, lesquelles considérez-vous « vraisemblables » ?
6,0 ; 6,1 ; 6,2 ; 6,3 ; 6,4 ; 6,5 ; 6,6 ; 6,7 ; 6,8 ; 6,9 ; 7,0

c) Supposons qu'en deux nuits, on observe en moyenne deux rongeurs par nuit. La valeur $\lambda = 5$ est-elle raisonnable ?

d) On compte le nombre de rongeurs en dix nuits. Soit $X_1, X_2, \ldots, X_n$ les n valeurs observées. Supposons qu'on estime λ par $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Supposons que $n = 10$ et $\lambda = 3$. Quelle est la probabilité d'une surestimation de plus de 10 % ?

e) Pour n assez grand, on peut considérer Y comme étant de loi approximativement normale. Montrer que la probabilité d'une surestimation de plus de 10 % est à peu près $1-\Phi[(0{,}1)\sqrt{n\lambda}]$

f) Supposons qu'en fait les valeurs exactes de $X_1; \ldots; X_n$ ne peuvent pas être observées : les traces laissées par le passage d'un rongeur permettent de conclure *qu'au moins* un rongeur est passé mais ne permettent pas de savoir *combien* sont passés. La seule information disponible est donc W = nombre de nuits (parmi n) durant lesquelles *aucun* rongeur n'est passé. Peut-on quand même estimer λ ? Sachant que $\hat{p} = W/n$ est une estimation raisonnable de la probabilité que X_i = 0, laquelle est égale à $e^{-\lambda}$, on estime $e^{-\lambda}$ par $\hat{p}$ et donc on peut estimer λ par $\hat{\lambda} = -\ln \hat{p}$. Si n est assez grand, W est à peu près de loi normale. Montrer que la probabilité d'une surestimation de plus de 10 % est alors à peu près égale à $P(\hat{\lambda} > 1{,}1\lambda) = \Phi\left[\frac{\sqrt{ne^{-\lambda}}(e^{-\lambda/10}-1)}{\sqrt{1-e^{-\lambda}}}\right]$.

g) Posez n = 100 et comparer $P(\bar{X} > 1{,}1\sqrt{\lambda})$ à $P(\hat{\lambda} > (1{,}1)\lambda)$ pour λ = 0,5 ; 1 ; 2 ; 3.

4.43 Soit $X_1; \ldots; X_n$ un échantillon aléatoire d'une population de loi $\mathcal{P}(\lambda)$.

a) Montrer que $\bar{X}$ et S^2 sont tous deux des estimateurs sans biais de λ.

b) Montrer que la loi conditionnelle du vecteur $\boldsymbol{X} = [X_1 ; X_2 ;\ldots; X_n]$ étant donné $\bar{X} = x/n$ est multinomiale de paramètres x et $[1/n; 1/n ;\ldots; 1/n]$.

c) Montrer que $E(X_i \mid \bar{X}) = \bar{X}$, $\mathrm{Var}(X_i \mid \bar{X}) = \frac{(n-1)\bar{X}}{n}$, et que $E(X_i^2 \mid \bar{x}) = \frac{(n-1)\bar{X}}{n} + \bar{X}^2$.

d) Montrer que $E(S^2 \mid \bar{X}) = \bar{X}$.

e) Montrer que $\mathrm{Var}(S^2) \geq \mathrm{Var}(\bar{X})$

4.44 Soit un échantillon aléatoire simple de taille n d'une population $\mathcal{N}(0 ; \sigma^2)$.

a) Montrer que l'estimateur du maximum de vraisemblance de σ^2 est $\frac{\sum_{i=1}^n X_i^2}{n}$. Montrer que cet estimateur est sans biais et déterminer sa variance.

b) On sait que l'estimateur $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}$ aussi est sans biais. Mais montrer qu'il est moins bon que celui présenté en a).

[*Rappel*: l'espérance d'une variable aléatoire de loi χ^2_ν est ν et sa variance est 2ν].

4.45 Afin d'estimer le nombre N de poissons dans une mare, vous en pêchez 30, vous les marquez, puis les remettez dans la mare. [Vous savez donc que la proportion de poissons marqués dans la marre est p = 30/N]. Ensuite vous repêchez 15 poissons, *avec* remise. Vous trouvez que parmi ces 15, 3 sont marqués. Déterminer l'estimateur du maximum de vraisemblance de N [*Suggestion* : commencez par estimer p].

4.46 Un appareil électrique fonctionne tant que les n piles qui y sont installées fonctionnent. Supposons que la durée de fonctionnement (en nombre d'heures) d'une pile est de loi exponentielle de moyenne β. Votre appareil a cessé de fonctionner exactement 12 heures après l'installation de n piles neuves (donc une des piles a cessé de fonctionner). À partir de cette seule information, estimez β par la méthode du maximum de vraisemblance. L'estimateur est-il sans biais ?

4.47 Voici la distribution observée du nombre d'accidents faits en une année par une classe de clients d'une compagnie d'assurance :

Nombre d'accidents	0	1	2	3	4	5
Nombre d'assurés	81 714	11 306	1618	250	40	7

a) Déterminer l'estimation du maximum de vraisemblance de p sous la supposition que le nombre d'accidents est de loi $\mathcal{B}(5; p)$.

b) Déterminer l'estimation du maximum de vraisemblance de λ sous la supposition que le nombre d'accidents est de loi $\mathcal{P}(\lambda)$.

c) Estimer la probabilité qu'un client n'ait aucun accident dans chacun des modèles ci-dessus. [Vous pouvez tenir pour acquis que si $\hat{p}$ est l'estimateur du maximum de vraisemblance de p, alors $(1-\hat{p})^n$ est l'estimateur du maximum de vraisemblance de $(1-p)^n$; de même, si $\hat{\lambda}$ est l'estimateur du maximum de vraisemblance de λ, alors $e^{-\hat{\lambda}}$ est l'estimateur du maximum de vraisemblance de $e^{-\lambda}$].

d) Déterminer l'estimation de la probabilité qu'un client n'ait aucun accident sans faire de supposition sur la distribution du nombre d'accidents.

4.48 Supposons que la durée de fonctionnement (en nombre d'heures) d'une *pile* est de loi exponentielle de moyenne β. Vous installez 4 piles dans un appareil électrique et deux autres piles dans un autre appareil. Les deux appareils ne fonctionnent que lorsque toutes les piles (4 dans le premier, 2 dans le deuxième) fonctionnent. Le premier appareil a cessé de fonctionner exactement 12 heures après que vous ayez installé 4 piles neuves (donc une des piles a cessé de fonctionner); le deuxième a cessé de fonctionner exactement 7 heures après l'installation des 2 piles. Estimer β par la méthode du maximum de vraisemblance.

4.49 Le nombre d'arrivées à un poste de péage dans l'intervalle de temps $[0; t]$ est une variable aléatoire de loi de Poisson de paramètre $\lambda = \alpha t$. Considérer les deux procédures suivantes pour estimer α.

Procédure 1 : Vous comptez le nombre d'arrivées X dans l'intervalle $[0; t]$ et vous estimez α par X/t.

Procédure 2 : Vous observez n arrivées successives et vous estimez α par $n/\Sigma Y_i$, Y_i étant le temps d'attente entre la $(i\text{-}1)^{e}$ et la i^{e} arrivées, $i = 2,\dots, n$, et Y_1 étant le temps d'attente avant la première arrivée.

Montrer que les deux estimateurs sont sans biais, et que leurs variances sont, respectivement, α/t et $\frac{(3n-2)\alpha^2}{(n-1)(n-2)}$. Déterminer une condition, en termes de n et t sous laquelle le 2^{e} estimateur est préférable au premier.

4.50 Le nombre de pièces défectueuses dans une boîte de k pièces est une variable de loi binomiale de paramètre p. Un fabricant vend n de ces boîtes avec la promesse de remplacer toute boîte dans laquelle au moins une pièce est défectueuse. Supposons que sur n boîtes vendues, Y n'ont aucune pièce défectueuse alors que les $n\text{-}Y$ en ont au moins une (on ne connaît rien de leur nombre, sinon qu'il est non nul). Peut-on quand même estimer p ? On sait que $\hat{\varphi} = Y/n$ est une estimation raisonnable de la probabilité qu'une boîte ne contienne aucune pièce défectueuse, donc une estimation de $\varphi = (1\text{-}p)^k$. Une estimation raisonnable de p est alors $\hat{p} = 1 - \hat{\varphi}^{1/k}$

a) Montrer que la probabilité de surestimer p par plus de 30 % est à peu près
$$\Phi\left[\sqrt{n}\,\frac{(1-1{,}3p)^k-(1-p)^k}{\sqrt{(1-p)^k[1-(1-p)^k]}}\right].$$

b) Évaluer cette probabilité pour $k = 4$, $n = 100$ et quelques valeurs de p.

c) Supposons qu'en fait on connaît le nombre de pièces défectueuses dans chaque boîte prélevée. On estime alors p par la proportion de pièces défectueuses parmi les nk pièces de l'échantillon. Déterminer approximativemen t la probabilité de surestimer p par plus de 30 % pour $k = 4$, $n = 100$ et les mêmes valeurs de p qu'en b).

4.51 Supposons que le nombre de fautes typographiques dans une page est une variable X de loi de Poisson de paramètre λ. Un vérificateur peut cependant ne pas découvrir toutes les fautes dans une page.

a) Soit Y le nombre de fautes *découvertes* par un vérificateur dans une page donnée. Supposons que $Y \sim \mathcal{B}(n\,;p)$, où n est le nombre réel de fautes dans la page. Montrer que Y est de loi de Poisson de paramètre $p\lambda$.

b) On demande au vérificateur de vérifier une page tirée au hasard. Soit Y le nombre de fautes qu'il a découvertes. À une autre occasion, on tire une autre page, mais cette fois-ci on s'assure (par une vérification très minutieuse) que toutes les fautes soient découvertes. Soit X le nombre de fautes découvertes. Montrer que $\hat{\lambda} = X$ et $\hat{p} = Y/X$ sont les estimateurs par les moments ainsi que les estimateurs à vraisemblance maximale.

c) Supposons maintenant qu'aux deux occasions on note le nombre réel de fautes X_1 et X_2 ainsi de le nombre de fautes découvertes par le vérificateur, Y_1 et Y_2. Montrez que les estimateurs du maximum de vraisemblance sont $\hat{\lambda} = \frac{X_1+X_2}{2}$ et $\hat{p} = \frac{Y_1+Y_2}{X_1+X_2}$.

d) Remarquez que l'estimateur en c) n'est pas défini si $X_1 + X_2 = 0$ et que $\mathrm{E}(\hat{p}) = \infty$, à moins qu'on décide arbitrairement de donner une valeur à $\hat{p}$ lorsque $X_1 + X_2 = 0$—disons $\hat{p} = 0$. On définirait alors $\hat{p}$ comme ceci : $\hat{p} = \begin{cases} \dfrac{Y_1+Y_2}{X_1+X_2} & \text{si } X_1+X_2>0 \\ 0 & \text{sinon} \end{cases}$. Montrer que le biais de $\hat{p}$ est $-pe^{-2\lambda}$.

4.52 Le nombre d'arrivées au quai d'un métro suit une loi de Poisson au taux de 2 arrivées/minute. Vous arrivez au quai à un moment aléatoire et trouvez 10 personnes sur le quai. Estimer le temps écoulé depuis le dernier départ.

4.53 Dans une population symétrique, la médiane et la moyenne coïncident, ce qui suggère que la médiane $\tilde{Y}$ d'un échantillon pourrait servir d'estimateur de la moyenne μ. Considérons un échantillon de taille $n = 2m+1$ d'une population de fonction de répartition F et de fonction de densité f. La fonction de densité de $\tilde{Y}$ est $g(\tilde{y}) = \dfrac{(2m+1)!}{m!m!} f(\tilde{y})[F(\tilde{y})]^m[1-F(\tilde{y})]^m$. Montrer que $\tilde{y}$ est sans biais. [Pour simplifier, définissez la variable $z = \tilde{y} - \mu$ et montrez que son espérance est nulle].

4.54 Soit $X_1, \ldots, X_n$ un échantillon d'une population de loi de Poisson de moyenne λ. Alors $\bar{X}$ et S^2 sont tous deux sans biais pour λ. Lequel est meilleur?

a) Soit $T = \sum_{i=1}^n X_i$. Montrer que la loi conditionnelle de $[X_1; \ldots; X_n]$ étant donnée $T = t$ est une loi multinomiale de paramètres t et probabilités toutes égales à $\dfrac{1}{n}$.

b) Montrer que $\mathrm{E}(\sum_{i=1}^n X_i^2 \mid \sum_{i=1}^n X_i = t) = \dfrac{(n-1)t+t^2}{n}$ et donc que $\mathrm{E}(S^2 \mid T = t) = \bar{X}$.

c) Montrer que $\mathrm{Var}(\bar{X}) \le \mathrm{Var}(S^2)$. [*Utilisez la formule* $\mathrm{Var}(\bar{X}) = \mathrm{Var}[E(\bar{X} \mid T = t)] + E[\mathrm{Var}(\bar{X} \mid T = t)]$]

d) Montrer que $\mathrm{Cov}(\bar{X}; S^2) = \mathrm{Var}(\bar{X})$. [Notez que $\mathrm{E}(\bar{X}S^2) = \mathrm{E}[\mathrm{E}(\bar{X}S^2 \mid T = t)]$.]

4.55 Soit n observations d'une population de densité

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \; ; x > 0, \theta > 0.$$

a) Montrer que $\bar{X}$ est sans biais pour β, et déterminer sa variance.

b) Soit $\theta = \beta^2$. Montrer que $\hat{\theta}_1 = \dfrac{1}{2n}\sum_{i=1}^n X_i^2$, $\hat{\theta}_2 = S^2 = \dfrac{\sum(X_i-\bar{X})^2}{n-1}$ et $\hat{\theta}_3 = \dfrac{n}{n+1}\bar{X}^2$ sont tous trois des estimateurs sans biais de θ. [Noter que pour une variable X de loi exponentielle, $\mathrm{E}(X^k) = k!\beta^k$]

c) Déterminer la variance de $\hat{\theta}_1$. [*Suggestion* : utiliser le fait que pour une variable X de loi exponentielle, $\mathrm{E}(X^k) = k!\beta^k$].

d) Déterminer la variance de $\hat{\theta}_3$.

e) Montrer que $\hat{\theta}_3$ est un meilleur estimateur que $\hat{\theta}_1$.

f) Montrer que $\dfrac{n-1}{n\bar{X}}$ est sans biais pour $1/\beta$.

4.56 [*Suite du numéro* précédent] Considérer l'estimateur $\hat{\theta}_3 = \dfrac{n\bar{X}^2}{n+1} = \dfrac{T^2}{n(n+1)}$, où $T = n\bar{X}$. Soit X l'une quelconque des variables $X_1, \ldots, X_n$.

a) Montrer que la densité conditionnelle de X étant donné $T = t$ est $f(x \mid t) = (n-1)\left(1-\frac{x}{t}\right)^{n-2}\frac{1}{t}$, $0 < x < t$.

b) Montrer que $\mathrm{E}(X \mid t) = \frac{t}{n}$ et que $\mathrm{E}(X^2 \mid t) = \frac{2t^2}{n(n+1)}$.

c) Montrer que $E(\hat{\theta}_1 \mid \hat{\theta}_3) = \hat{\theta}_3$, où $\hat{\theta}_1 = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$. [Puisque $\hat{\theta}_3$ est une fonction bijective de T, il revient au même de montrer que $E(\hat{\theta}_1 \mid T = t) = \frac{t^2}{n(n+1)}$, ce qui est moins encombrant.]

d) Montrer que $E(\hat{\theta}_2 \mid \hat{\theta}_3) = \hat{\theta}_3$, où $\hat{\theta}_2 = S^2$.

e) Montrer que $\mathrm{Var}(\hat{\theta}_1) \geq \mathrm{Var}(\hat{\theta}_3)$ . .

f) Montrer que $\mathrm{Var}(\hat{\theta}_2) \geq \mathrm{Var}(\hat{\theta}_3)$.

g) Montrer que $\mathrm{Cov}(\hat{\theta}_1;\hat{\theta}_3) = \mathrm{Var}(\hat{\theta}_3)$.

h) Montrer que $\mathrm{Cov}(\hat{\theta}_2;\hat{\theta}_3) = \mathrm{Var}(\hat{\theta}_3)$.

4.57 Une observation X de loi binomiale négative de paramètre r (connu) et p. On sait que l'estimateur de p par la moyenne ainsi que par le maximum de vraisemblance est $\hat{p} = \frac{r}{X}$. Il existe un autre critère qui mène à un estimateur: l'estimateur est la valeur de p l'estimateur qui minimise la somme des carrés Q $= \sum_{i=1}^{r}[X_i - \mathrm{E}(X_i)]^2 = \sum_{i=1}^{r}[X_i - 1/p]^2$. Montrer que cette approche aussi mène à $\hat{p} = \frac{r}{X}$.

4.58 D'une cargaison d'oranges, de deux sources, A et B, on tire un échantillon de 100 oranges. La source des oranges ne peut pas être identifiée visuellement, on ne connaît pas la proportion d'oranges de chaque source. On s'intéresse à estimer la proportion d'oranges de source A dans la cargaison. La teneur en sucre X_1 et X_2 des deux sources sont des variables dont les paramètres sont bien connus. Soit μ_1 et σ_1 la moyenne et l'écart type de la teneur en sucre des oranges de source A et μ_2 et σ_2 celles des oranges de source 2.

Soit Z_1, Z_2, … , Z_n la teneur en sucre des n oranges. Donc $Z_i = X_{1i}$ si l'orange i est de source A et $X_{2\mathrm{i}}$ si l'orange i est de source B.

a) Soit Z la teneur en sucre d'une orange tirée au hasard dans le lot

i) Montrer que $\mathrm{E}(Z) = p\mu_1 + q\mu_2$, où p est la proportion d'oranges de source A dans le lot et $q = 1 - p$.

ii) Soit $\sigma^2 = \mathrm{Var}(Z)$. Montrer que $\sigma^2 = pq(\mu_1 - \mu_2)^2 + p\sigma_1^2 + q\sigma_2^2$

iii) Montrer que $\hat{p} = \frac{\bar{Z} - \mu_2}{\mu_1 - \mu_2}$ est un estimateur sans biais de p.

iv) Montrer que $\hat{V}_1 = \frac{\hat{p}\sigma_1^2 + \hat{q}\sigma_2^2 + \hat{p}\hat{q}(\mu_1 - \mu_2)^2}{(n-1)(\mu_1 - \mu_2)^2}$ est un estimateur sans biais de la variance de $\hat{p}$ ($\hat{q} = 1 - \hat{p}$).

v) Montrer que $\hat{V}_2 = \frac{S^2}{n(\mu_1 - \mu_2)^2}$ est un autre estimateur sans biais de la variance de $\hat{p}$ [$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$].

b) Pour une population données, on a

Source A: $\mu_1 = 9$ g et $\sigma_1 = 1{,}5$ g ; Source B: $\mu_2 = 12$ g et $\sigma_2 = 2{,}5$.

Un échantillon de 100 oranges donne $\bar{Z}$ = 11,1835 et S = 2,8215. Déterminer $\hat{p}$ et $\hat{V}_2$.

4.59 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de loi uniforme sur [0 ; β].

Soit $X = X_1$ et $Y = X_{(n)}$. Soit $\hat{\beta}_1 = 2X$ et $\hat{\beta}_2 = \frac{n+1}{n}Y$.

a) Montrer que $\hat{\beta}_1$ et $\hat{\beta}_2$ sont tous deux sans biais.

b) Montrer que $E(\hat{\beta}_1 \mid Y) = \hat{\beta}_2$ (donc que $E(\hat{\beta}_1 \mid \hat{\beta}_2) = \hat{\beta}_2$).

c) Concluez que $\mathrm{Var}(\hat{\beta}_1) \geq \mathrm{Var}(\hat{\beta}_2)$ du fait que $\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}[E(\hat{\beta}_1 \mid \hat{\beta}_2)] + E[\mathrm{Var}(\hat{\beta}_1 \mid \hat{\beta}_2)]$.

4.60 Soit $X_1; \ldots; X_n$ un échantillon aléatoire simple d'une population de fonction de densité f et de fonction de répartition F.

a) Soit Y la n^{e} statistique d'ordre et $Y_1, Y_2, \ldots, Y_{n-1}$ les $n - 1$ autres valeurs de l'échantillon. Montrer que la distribution conjointe de $\boldsymbol{Y} = [Y_1 ; Y_2 ; \ldots ; Y_{n-1}, Y]$ est $nf(y)\prod_{i=1}^{n-1} f(y_i)$.

b) Montrer que la fonction de densité conditionnelle de $[Y_1, Y_2, \ldots, Y_{n-1}]$ étant donné $Y = y$ est

$$\frac{\prod_{i=1}^{n-1} f(y_i)}{[F(y)]^{n-1}} = \prod_{i=1}^{n-1}\left(\frac{f(y_i)}{F(y)}\right)$$

c) Montrer que si la population est de loi $\mathcal{U}(0 ; \beta)$, alors la fonction de densité conditionnelle de $[Y_1, Y_2, \ldots, Y_{n-1}]$ étant donné $Y = y$ est $\frac{1}{y^{n-1}}$. En d'autres termes, les Y_i sont conditionnellement un échantillon aléatoire simple d'une population de loi $\mathcal{U}(0 ; y)$. On remarque ici que la distribution conditionnelle des observations étant donné leur valeur maximale est indépendante du paramètre. En termes concrets, cela signifie que si l'on connaît le maximum, le reste des observations de fournit plus aucune information additionnelle concernant le paramètre.

d) Soit $\bar{X}$ la moyenne des observations et $\hat{\beta}_1 = 2\bar{X}$ et $\hat{\beta}_2 = \frac{n+1}{n}Y$. Montrer que $E(\hat{\beta}_1 \mid Y) = \hat{\beta}_2$ (et donc que $E(\hat{\beta}_1 \mid \hat{\beta}_2) = \hat{\beta}_2$).

4.61 On suppose que parmi les nouveaux arrivés dans le pays, une certaine proportion p a été admise avec de faux papiers. Afin de se faire une idée de la valeur de p, on tire des individus au hasard jusqu'à ce qu'on tombe sur une personne avec de faux papiers. On compte alors le nombre X d'unités qu'il a fallu tirer. On ne peut pas estimer p sans biais, mais on peut estimer sa réciproque $\theta = 1/p$, une information qui n'est pas dénuée de sens : $\theta - 1$ est le nombre d'immigrants légitimes pour chaque immigrant illégal. Soit X le nombre d'essais effectués au moment où on tombe sur un premier immigrant illégal. Considérons l'estimateur $\hat{\theta} = X$.

a) Montrer que $\hat{\theta}$ est un estimateur sans biais de θ.

b) Montrer que $\mathrm{Var}(\hat{\theta}) = (1-p)/p^2$.

c) Montrer que $\hat{V} = \frac{X(X-1)}{2}$ est un estimateur sans biais de $\mathrm{Var}(\hat{\theta})$.