STT1000

Chapitre 5 Intervalles de confiance

Solutions

Intervalle de confiance pour une moyenne

5.1 Dans sa chasse aux insectes volants, la chauve-souris émet des sons de haute fréquence dont elle reçoit ensuite l'écho. Les pulsations, normalement émises à des intervalles de 50 à 100 millièmes de seconde, ralentissent brusquement dès que la chauve-souris perçoit l'insecte. Profitant de ce fait, des expérimentateurs ont étudié la distance à laquelle la chauve-souris perçoit l'insecte. À l'aide d'une caméra et d'un appareil mesurant le rythme des pulsations ils ont pu déterminer la distance qui séparait la chauve-souris de l'insecte au moment précis où l'insecte a été détecté. Cette expérience a été répétée 11 fois. Les distances (en centimètres) sont les suivantes :

62 ; 52 ; 68 ; 23 ; 34 ; 45 ; 27 ; 42 ; 83 ; 56 ; 40

Déterminer un intervalle de confiance à 95 % pour μ, la distance moyenne nécessaire à la détection d'un insecte par la chauve-souris.

$n = 11$; $\bar{X} = 48{,}36364$; $S = 18{,}08465$; $\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}} = 5{,}452727$; point critique (à 95 %) = 2,228.

Marge d'erreur : $12{,}14843\hat{\sigma}_{\bar{X}} = 11{,}805$. Intervalle de confiance : [36,21 ; 60,51].

5.2 [*Données du tableau* A05] Une vieille règle affirme que la température normale (donc moyenne) du corps est de 98,6 °F (des études récentes, cependant, fixent le chiffre à 98,2). Déterminez, à partir des données du tableau, un intervalle de confiance pour la température moyenne du corps humain. Vos données vous permettent-elles de conclure que l'une ou l'autre des deux études donne des résultats incorrects ? [Données de l'échantillon: $\bar{X} = 98{,}24923$; $S = 0{,}7331832$; $n = 130$].

Point critique à 95 % : 1,978524 ; marge d'erreur : 0,1272279 ; intervalle de confiance : $98{,}122 \leq \mu \leq 98{,}376$.
Selon ces données, la vieille règle n'est pas bonne puisque la moyenne proposée de 98,6 °C ne se situe pas dans l'ensemble des valeurs vraisemblables de μ.

Intervalle de confiance pour une proportion

5.3 L'administration d'un service d'urgences souhaite estimer la proportion p de ses patients dont la résidence se situe dans le secteur administratif de l'hôpital. Dans un échantillon de 200 patients, on en trouve 60 dans cette catégorie. Déterminer un intervalle de confiance à 95 % pour p.

$\hat{p} = 0{,}3$; $\hat{\sigma}_{\hat{p}} = 0{,}0324037$; marge d'erreur (à 95 %) : 0,06351126 ; intervalle de confiance : [0,236 ; 0,364]

5.4 [*Données du tableau* A03] Déterminez un intervalle de confiance pour la proportion des gens dont le score au test verbal (V) est supérieur au score dans le test global (G). Le nombre de cas où le test verbal est supérieur au score global est 9, parmi 28.

$\hat{p} = \frac{9}{28} = 0{,}3214$. L'écart-type de $\hat{p}$ est estimé par $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0{,}088$. Un intervalle de confiance à 95 % est donc donné par [$\hat{p} - 1{,}96\hat{\sigma}_{\hat{p}}$; $\hat{p} + 1{,}96\hat{\sigma}_{\hat{p}}$] = [14,84 % ; 49,44 %].

5.5 D'une population de comptes de dépenses on tire un échantillon de 120 comptes pour estimer la proportion p des comptes qui comportent un voyage en dehors de la ville. Soit X le nombre de comptes dans l'échantillon qui comportent un voyage en dehors de la ville. Si $X = 28$, déterminer un intervalle de confiance à 95% pour p.

$\hat{p} = 0{,}2333$; $\hat{\sigma}_{\hat{p}} = 0{,}0386$; marge d'erreur (à 95 %) : 0,0757. Intervalle de confiance : [0,158 ; 0,309].

5.6 Montrez qu'un intervalle construit par la formule $\hat{p} - 1{,}88\,\hat{\sigma}_{\hat{p}} \leq p \leq \hat{p} + 2{,}05\,\hat{\sigma}_{\hat{p}}$ est un intervalle de confiance à 95% (approximativement) [vous supposerez que $\frac{\hat{p}-p}{\hat{\sigma}_{\hat{p}}} \sim N(0\,;1)$].

L'intervalle présenté est équivalent à l'intervalle $-2,05 \leq \frac{\hat{p}-p}{\hat{\sigma}_{\hat{p}}} \leq 1,88$ et si on suppose que la variable centrée-réduite au centre est $\mathcal{N}(0 ; 1)$, alors $P\left(\hat{p}-1,88\hat{\sigma}_{\hat{p}} \leq p \leq \hat{p}+2,05\hat{\sigma}_{\hat{p}}\right) = P\left(-2,05 \leq \frac{\hat{p}-p}{\hat{\sigma}_{\hat{p}}} \leq 1,88\right) = 0,95$, ce qui montre que l'intervalle $\left[\hat{p}-1,88\hat{\sigma}_{\hat{p}} ; \hat{p}+2,05\hat{\sigma}_{\hat{p}}\right]$ a une probabilité de recouvrir p de 95 % (à peu près).

Intervalle de confiance pour une variance

5.7 Au numéro 5.1, déterminer un intervalle de confiance à 85 % pour l'écart-type σ des distances.
$n = 11$; $S = 18,08465$; $(n-1)S^2 = 3270,545$; points critiques : 3,246973 et 20,48318.
Intervalle de confiance = [159,6698 ; 1007.2599

5.8 Au numéro 5.1, montrer que tous les intervalles de confiance dans le tableau suivant sont de niveau 85 %.

Intervalle de confiance		Largeur de l'intervalle
Limite Inférieure	Limite supérieure	
221,17	1278,45	1057,28
215,19	1007,26	792,07
204,57	830,02	625,45
192,71	735,64	542,93
140,92	601,32	460,41
0,00	587,17	587,17
225,02	∞	∞

5.9 Au numéro 5.2, soit σ, σ_f et σ_h l'écart-type des températures du corps humain, des femmes et des hommes, respectivement. Déterminer un intervalle de confiance à 90 % pour σ, pour σ_f et pour σ_h. Les écarts-types échantillonnaux sont $S = 0,7331832$, $S_f = 0,7434878$ et $S_h = 0,6987558$. L'échantillon est constitué de 65 femmes et 65 hommes.

Femmes

n	S	$\chi^2_{0,05}$	$\chi^2_{0,95}$	$(n-1)S^2$	Intervalle de confiance
65	0,74349	83,67526	46,59491	35,37754	0,4228 ; 0,7593

Hommes

n	S	$\chi^2_{0,05}$	$\chi^2_{0,95}$	$(n-1)S^2$	Intervalle de confiance
65	0,69876	83,67526	46,59491	31,24862	0,3735 ; 0,6706

Tous

n	S	$\chi^2_{0,05}$	$\chi^2_{0,95}$	$(n-1)S^2$	Intervalle de confiance
130	0,73318	156,5075	103,7646	69,34492	0,4431 ; 0,6683

5.10 [*Suite du numéro précédent*] Une façon de décider si $\sigma_f = \sigma_h$ consiste à déterminer un intervalle de confiance pour chacun des deux paramètres et de conclure que $\sigma_f \neq \sigma_h$ si les deux intervalles ne se recoupent pas. Est-ce le cas ? [*Ce n'est pas la façon optimale de procéder. Une meilleure façon de le faire est présentée au chapitre* 6].
Avec cette règle, on ne peut pas conclure qu'il y a une réelle différence entre la dispersion des femmes et la dispersion des hommes. Cette procédure n'est pas optimale dans le sens qu'elle manque de puissance.

Exercices divers

5.11 [*Données du tableau* A03] Considérez les sujets comme un échantillon d'une population de jeunes nord-américains de race blanche. Déterminez un intervalle de confiance pour μ, la taille moyenne de la population ; pour μ_f, la taille moyenne des femmes de la population ; et pour μ_h, la taille moyenne des hommes de cette population. La marge d'erreur devrait être à peu près inversement proportionnelle à la racine carrée de la taille de l'échantillon. Est-ce le cas ici ? Discutez.
Voici les moyennes et écarts-types des trois groupes (le groupe entier, les femmes, les hommes) :

	n	Moyenne	S	Pt critique (95%)	Intervalle de confiance	Marge d'erreur
Groupe entier	28	68,161	3,812	2,052	66,68 ; 69,64	1,48
Femmes	16	65,96	2,388	2,131	64,68 ; 67,23	1,27
Hommes	12	71,1	3,377	2,201	68,95 ; 73,25	2,15

Pour le groupe entier, nous avons un grand échantillon (n = 28), ce qui devrait réduire la marge d'erreur. Sauf que, la population étant hétérogène (hommes et femmes), la dispersion est grande (S = 3,812), ce qui fait croître la marge d'erreur —un accroissement atténué en partie seulement par la (relativement) grande taille de l'échantillon.
Pour les femmes, la marge d'erreur relativement étroite est attribuable au fait que la dispersion chez les femmes est petite (S = 2,388), un phénomène qu'on observe dans plusieurs comparaisons entre femmes et hommes.
La grande marge d'erreur pour les hommes reflète en partie la petite taille de l'échantillon (12) et en partie la grande dispersion de la population. (S = 3,377).

5.12 [*Données du tableau* A09] Considérez la variable D = [(B1+B2 - (A1+A2)]/2 comme mesure du gain réalisé par l'apprentissage. Déterminez un intervalle de confiance pour le gain moyen dû à l'apprentissage, pour chacune des trois méthodes.

Voici les valeurs de la variable D :

```
 1.0   1.5  -2.5  -2.5  -1.0  -5.5  -2.5  -4.5   0.0  -1.0  -2.0  -1.5  -3.5   1.0  -2.0  -0.5  -3.5
-3.5  -2.5  -3.5  -0.5   0.0   2.0  -1.0   0.0   0.5  -1.5  -1.0   2.0   1.5  -0.5  -1.5   0.0  -0.5
 2.0   0.5   1.0   4.5   2.0  -1.5   2.5   0.5   1.5   1.0   2.5  -0.5   2.0  -0.5  -0.5   1.0   2.0
 3.0   2.0   2.0   0.0  -2.0   0.5  -3.5   3.0   1.5   2.0   1.5   2.0   2.0   3.0  -0.5.
```

Voici les données de base et les intervalles de confiance pour chacun des sous-échantillons ainsi que pour l'échantillon complet :

	n	Moyenne	S	Pt critique	Limite inférieure	Limite supérieure
Groupe entier	66	-0,053	2,092	1,9971	-0,567	0,461
Type 1	22	-1,773	1,856	2,080	-2,596	-0,950
Type 2	22	0,591	1,563	2,080	-0,102	1,284
Type 3	22	1,023	1,708	2,080	0,266	1,780

5.13 Une institution scolaire possède 480 salles de cours. Dans un échantillon de 40 salles, on trouve que 5 salles ont un rétroprojecteur endommagé par des vandales. Déterminer un intervalle de confiance à 95 % pour le nombre M de rétroprojecteurs endommagés.

$\hat{p}$ = 0,125 ; $\hat{\sigma}_{\hat{p}}$ = 0,05229125 ; marge d'erreur (à 95 %) pour l'estimation de p = 0,10249.

Intervalle de confiance pour p : [0,02250915 ; 0,22749085].
Si on peut affirmer que p se trouve dans cet intervalle, alors on peut affirmer que le *nombre* de salles avec rétro défectueux est dans l'intervalle : [480×0,02250915 ; 480×0,22749085] = [11 ; 109]. Pour être sûr, on affirmera que le nombre en question est de 10 à 110 (y compris les bornes)
Remarque Toutes les formules d'intervalle de confiance données dans ce cours sont basées sur la supposition que la population est infinie. Dans la plupart des cas cette supposition est anodine, la taille de la population étant généralement très grande (par rapport à la population). Dans le cas présent, la population est relativement petite et un certain ajustement s'impose. Sans donner les détails de cet ajustement, nous pouvons en indiquer les résultats. L'écart-type de l'estimateur est estimé à 0,05070286 (au lieu de 0,05229125) et l'intervalle de confiance pour le nombre de salles avec rétro endommagé est [12 ; 108] au lieu de [11 ; 109]. La différence n'est pas négligeable, mais certainement pas aussi importante qu'on le croirait à première vue.

5.14 Une compagnie prélève un échantillon de 50 chèques parmi les 2 500 reçus en une journée donnée. On suit le parcours des chèques jusqu'au moment de leur dépôt dans le compte de la compagnie. On constate que 18 des 50 chèques ont mis plus de 5 jours à être déposés.

a) Déterminer un intervalle de confiance à 95% pour la proportion p de chèques dont le délai (entre la réception et le dépôt) excède 5 jours ;
$\hat{p}$ = 0,36 ; $\hat{\sigma}_{\hat{p}}$ = 0,06788225 ; marge d'erreur (à 95 %) = 0,1330468 ; intervalle de confiance : [0,22695 ; 0,49305].

b) Déterminez un intervalle de confiance à 95% pour le nombre de chèques dont le délai excède 5 jours.
Il suffit de multiplier les bornes de l'intervalle pour p pour avoir celle du nombre total : [567 ; 1233].

c) Supposons qu'on veuille faire un échantillonnage sur les chèques de l'année entière (au nombre de 650 000). À un niveau de 95%, quelle est la taille de l'échantillon qu'il faudrait prélever dans les conditions suivantes (vous prendrez pour p l'estimation que vous obtenez avec l'échantillon que vous venez de prélever) ?
Pour avoir une marge de m, c'est-à-dire, s'il faut que $1{,}96\sqrt{\frac{p(1-p)}{n}} = m$, alors il faut que $\sqrt{n} = \frac{1{,}96\sqrt{p(1-p)}}{m} \Rightarrow n = \frac{p(1-p)}{[m/1{,}96]^2}$.
Nous estimerons p par la proportion de l'échantillon en cours, soit p = 0,36.
   i) si on accepte une marge de 2% dans l'estimation de la proportion ;
   Ici $m = 0{,}02 \Rightarrow n = 2213$
   ii) si on accepte une marge d'erreur *relative* (voir le numéro précédent) de 5% de la proportion réelle ;
   $m = 0{,}05p = 0{,}018 \Rightarrow n = 2732$

iii) si on accepte une marge d'erreur de 10 000 chèques dans l'estimation du nombre de chèques qui accusent un délai de plus de 5 jours.

Une marge d'erreur de 10 000 chèques est une marge d'erreur de 10 000/650000 dans l'estimation de la proportion. Donc m = 10 000/650000 $\Rightarrow n$ = 3740

Exercices théoriques

5.15 Quelle est la taille de l'échantillon qu'on devrait prélever pour estimer une proportion p avec une probabilité d'au moins 95 % de ne pas se tromper de plus d'un point de pourcentage? On ne connaît pas la valeur de p mais on peut être sûr qu'elle n'est pas supérieure à 15%. [Servez-vous du fait que la fonction $p(1-p)$ est croissante sur l'intervalle (0 ; 1/2).]

Nous voulons que $P(|\hat{p}-p| \leq 0{,}01) \geq 0{,}95 \Rightarrow P\left(\frac{|\hat{p}-p|}{\sigma_{\hat{p}}} \leq \frac{0{,}01}{\sigma_{\hat{p}}}\right) \geq 0{,}95 \Rightarrow \frac{0{,}01}{\sigma_{\hat{p}}} \geq 1{,}96 \Rightarrow \sigma_{\hat{p}} \leq \frac{0{,}01}{1{,}96} \Rightarrow \sqrt{\frac{p(1-p)}{n}} \leq \frac{0{,}01}{1{,}96} \Rightarrow$

$\sqrt{n} \geq \frac{\sqrt{p(1-p)}}{0{,}01/1{,}96} \Rightarrow n \geq \frac{p(1-p)}{(0{,}01/1{,}96)^2}$. Le numérateur $p(1-p)$ est inconnu, mais on sait qu'il est $\leq$ (0,15)(0,85), puisque la fonction $p(1-p)$ est croissante sur [0 ; ½] et $p \leq 0{,}15$ par hypothèse. Donc la valeur maximale de $\frac{p(1-p)}{(0{,}01/1{,}96)^2}$ est $\frac{(0{,}15)(0{,}85)}{(0{,}01/1{,}96)^2}$ = 4898. Cette taille énorme est due à l'extrême exigence de précision. Si on tolérait plutôt une précision de 5 % (au lieu de 1 %), un échantillon de taille 196 suffirait.

5.16 Considérons le problème qui consiste à déterminer la taille de l'échantillon qu'on devrait prélever pour qu'avec une probabilité de 95% l'erreur *relative* dans l'estimation de p ne dépasse pas 20 % [L'erreur relative est l'erreur exprimée comme proportion de p : erreur relative = $|\hat{p}-p|/p$]. Supposer que la meilleure estimation disponible de p est p = 0,10 et estimer la taille de l'échantillon qu'il faut prélever.

Ce qui est demandé, c'est que $P\left(\frac{|\hat{p}-p|}{p} \leq 0{,}2\right)$ = 0,95, ce qui entraîne que $\frac{0{,}2p}{\sigma_{\hat{p}}}$ =1,96 Þ $n = \left(\frac{1{,}96}{0{,}2}\right)^2 \frac{1-p}{p}$. Si on estime p par 0,1, on a n = 864.

5.17 Soit X_1 et X_2 un échantillon aléatoire de taille n = 2 d'une population $\mathcal{N}(\mu ; \sigma^2)$. Déterminer un intervalle de confiance pour σ^2. [Considérer la distribution de $\frac{(X_1-X_2)^2}{2\sigma^2}$.]

X_1-X_2 est de loi normale de moyenne 0 et de variance $2\sigma^2$. Donc $\frac{X_1-X_2}{\sqrt{2\sigma^2}}$ est de loi $\mathcal{N}(0 ; 1)$ et $\frac{(X_1-X_2)^2}{2\sigma^2} \sim \chi^2_1$. Donc

$P\left(0{,}001 \leq \frac{(X_1-X_2)^2}{2\sigma^2} \leq 5{,}024\right)$ = 0,05, ce qui donne l'intervalle suivant : $\frac{(X_1-X_2)^2}{2(5{,}024)} \leq \sigma^2 \leq \frac{(X_1-X_2)^2}{2(0{,}001)}$.

5.18 On sait que si $X \sim \chi^2_\nu$ et si ν est grand, alors la variable $\sqrt{2X}-\sqrt{2\nu}$ suit approximativement une loi $\mathcal{N}(0 ; 1)$. Servez-vous de ce résultat pour montrer que si n est grand, alors S suit à peu près une loi $\mathcal{N}(\sigma ; \sigma^2/2(n-1))$. Déterminer un intervalle de confiance à 95% pour σ, sachant qu'un échantillon de taille 250 a donné S = 39.

On pose $X = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. $\sqrt{2X}-\sqrt{2(n-1)} = \sqrt{2(n-1)}\left(\frac{S}{\sigma}-1\right) \sim \mathcal{N}(0\,;1) \Rightarrow \sigma\left(\frac{S}{\sigma}-1\right) = S-\sigma \sim \mathcal{N}\left(0; \frac{\sigma^2}{2(n-1)}\right) \Rightarrow$

$S \sim \mathcal{N}\left(\sigma; \frac{\sigma^2}{2(n-1)}\right)$, à peu près. Pour déterminer un intervalle de confiance à 95 %, on écrit

$\mathrm{P}\left[-1{,}96 \leq \sqrt{2(n-1)}\left(\frac{S}{\sigma}-1\right) \leq 1{,}96\right] = 0{,}95 \Rightarrow P\left(\frac{S}{1+1{,}96/\sqrt{2(n-1)}} \leq \sigma \leq \frac{S}{1-1{,}96/\sqrt{2(n-1)}}\right)$ = 0,95. L'intervalle est donc

$\left[\frac{S}{1+1{,}96/\sqrt{2(n-1)}} ; \frac{S}{1-1{,}96/\sqrt{2(n-1)}}\right] = \left[\frac{39}{1+1{,}96/\sqrt{2(249)}} ; \frac{39}{1-1{,}96/\sqrt{2(249)}}\right]$.

5.19 On peut définir la loi de *Student* en ces termes: soit X et Y deux variables aléatoires *indépendantes*, $X \sim N(0\,;1)$, $Y \sim \chi^2_\nu$. Alors la variable $T = \dfrac{X^2}{\sqrt{Y^2/\nu}} \sim t_\nu$. Montrer que, si $\nu > 2$, $E[T] = 0$, et $\mathrm{Var}[T] = E[T^2]$. [*Suggestion*: Noter que $E(T) = E[X/(Y/\nu)] = \nu\, E[X\,Y^{-1}] = \nu\, E[X]\, E[Y^{-1}]$ où la dernière égalité découle du fait que l'espérance d'un produit de variables indépendantes est égale au produit de leurs espérances; et $\mathrm{Var}[T] = \nu\, E[X^2]\, E[Y^{-1}]$]

Sachant que Y est de loi χ^2_ν, on a $E(Y^{-1}) = \dfrac{1}{\Gamma(\nu/2)2^{\nu/2}}\displaystyle\int_0^\infty y y^{\nu/2-1}e^{-y/2}dy = \dfrac{\Gamma(\nu/2-1)2^{\nu/2-1}}{\Gamma(\nu/2)2^{\nu/2}} = \dfrac{\Gamma(\nu/2-1)2^{-1}}{(\nu/2-1)\Gamma(\nu/2-1)} = \dfrac{1}{\nu-2}$.

Puisque $E[X^2] = 1$, $\mathrm{Var}[T] = \nu\, E[X^2]\, E[Y^{-1}] = \nu\, E[Y^{-1}] = \nu/(\nu-2)$.

5.20 La durée d'une certaine pièce électronique est une variable de loi exponentielle de moyenne β. Un échantillon de 4 pièces donne les durées suivantes (en milliers d'heures) : 2,4 ; 3,0 ; 8,4 ; 11,0. Déterminer un intervalle de confiance à 80 % pour β [utiliser $2\Sigma X_i/\beta$ comme pivot].

$\dfrac{2\sum X_i}{\beta} \sim \chi^2_8$. Donc $P\left[\chi^2_{8,\alpha/2} \le \dfrac{2\sum X_i}{\beta} \le \chi^2_{8,1-\alpha/2}\right] = 1-\alpha \Rightarrow P\left[\dfrac{2\sum X_i}{\chi^2_{8,1-\alpha/2}} \le \beta \le \dfrac{2\sum X_i}{\chi^2_{8,\alpha/2}}\right] = 1-\alpha$. Un intervalle de confiance est donc donné par $\dfrac{2\sum X_i}{\chi^2_{8,1-\alpha/2}} \le \beta \le \dfrac{2\sum X_i}{\chi^2_{8,\alpha/2}} \Leftrightarrow \dfrac{49{,}6}{13{,}3616} \le \beta \le \dfrac{49{,}6}{34{,}4895} \Leftrightarrow 3{,}71 \le \beta \le 14{,}21$.

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

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

a) Considérer les variables $Y_i = -\ln X_i$. Montrer que la fonction de densité de Y_i est $g(y) = \dfrac{1}{\theta}e^{-y/\theta}$.

La fonction de répartition de X est $F(x) = x^{1/\theta}$, $x < 1$. La fonction de répartition de Y est $G(y) = P(Y \le y) = P(X \ge e^{-y}] = 1 - F(e^{-y}) = 1-e^{-y/\theta}$, $y > 0 \Rightarrow g(y) = \dfrac{1}{\theta}e^{-y/\theta}$, $y > 0$.

b) Déterminer un intervalle de confiance pour θ [Considérez la loi des $Z_i = 2Y_i/\theta$ puis celle de leur moyenne $\bar{Z}$].

Puisque Y est de loi exponentielle de moyenne θ, sa fonction génératrice des moments est $(1-\theta t)^{-1}$. Donc celle de Z est $(1-2t)^{-1}$, la fonction génératrice des moments de Z est $(1-2t)^{-1}$, qui est celle d'une variable de loi χ^2 à 2 degrés de liberté$\Rightarrow \Sigma Z_i = 2\Sigma Y_i/\theta = -2\Sigma \ln X_i/\theta$ est de loi χ^2 à $2n$ degrés de liberté. $P\left(\chi^2_{2n;1-\alpha/2} \le \dfrac{-2\sum_{i=1}^n \ln X_i}{\theta} \le \chi^2_{2n;\alpha/2}\right) = 1-\alpha$, ce qui donne pour intervalle

$$\frac{-2\sum_{i=1}^n \ln X_i}{\chi^2_{2n;\alpha/2}} \le \theta \le \frac{-2\sum_{i=1}^n \ln X_i}{\chi^2_{2n;1-\alpha/2}}.$$

5.22 Soit une observation d'une population de densité

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

a) Déterminer la fonction de répartition de $Y = X^{1/\theta}$.

$X^{1/\theta}$ est la fonction de répartition de X et donc Y est de loi uniforme sur [0 ; 1].

b) Déterminer un intervalle de confiance à 95% pour θ à partir d'une seule observation tirée de cette population. Un intervalle possible est basé sur le fait que $P(\alpha/2 \le X^{1/\theta} \le 1-\alpha/2) = 1-\alpha \Leftrightarrow \ln X/\ln(\alpha/2) \le \theta \le \ln X/\ln(1-\alpha/2)$.

5.23 Soit n observations d'une population de densité

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

Déterminer un intervalle de confiance pour θ en partant du fait que $Y = \frac{2}{\theta}\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$. $\frac{2\sum X_i}{\chi^2_{2n;\alpha/2}} \le \theta \le \frac{2\sum X_i}{\chi^2_{2n;1-\alpha/2}}$

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

$$f(x) = \theta x^{-(\theta+1)},\ x > 1,\ \theta > 0.$$

a) Montrer que la variable $Y_i = \theta\ln(X_i)$ suit une loi exponentielle de paramètre 1.
La fonction de répartition de X est $F(x) = 1\text{-}x^{-\theta}$, $x > 1$.
La fonction de répartition de Y est $G(y) = P(\theta\ln X \le y) = P(X \le e^{y/\theta}) = 1 - e^{-y}$; c'est bien la fonction de répartition d'une variable de loi exponentielle de paramètre 1.

b) Déterminez la distribution de $Z = \sum Y_i$ et déterminez un intervalle de confiance pour θ à partir d'un échantillon de n observations X_1 ; X_2 ; … ; X_n.
On utilise ici la même démarche qu'au numéro précédent.

c) Montrer que $\frac{n-1}{\sum_{i=1}^{n}\ln X_i}$ est un estimateur sans biais de θ.
Si $U_i = \theta\ln X_i$ est de loi exponentielle de paramètre 1, alors $U = \sum U_i = \sum\theta\ln X_i$ est de loi gamma de paramètres n et 1. Alors
$E(U) = E\left(\frac{1}{\theta\sum\ln X_i}\right) = \frac{1}{n-1} \Rightarrow E\left(\frac{n-1}{\sum\ln X_i}\right) = \theta$.

5.25 Sur une certaine ligne de métro les trains arrivent régulièrement à toutes les θ minutes. Vous arrivez au quai à un moment aléatoire et vous attendez le prochain métro. Soit X votre temps d'attente.

a) Utiliser le fait que $Y = X/\theta$ est de loi uniforme sur [0 ; 1] pour montrer que l'intervalle $\left[\frac{X}{1-\alpha/2} \le \theta \le \frac{X}{\alpha/2}\right]$ est un intervalle de confiance à 100(1-α) % pour θ.
La probabilité que l'intervalle contienne le paramètre est
$P\left[\frac{X}{1-\alpha/2} \le \theta \le \frac{X}{\alpha/2}\right] = P\left[\alpha/2 \le \frac{X}{\theta} \le 1-\alpha/2\right] = 1-\alpha/2-\alpha/2 = 1-\alpha$

b) Montrer que pour α fixe et 0 < β ≤ α, l'intervalle $\left[\frac{X}{1-(\alpha-\beta)} \le \theta \le \frac{X}{\beta}\right]$ est un intervalle de confiance de niveau 1-α. Quelle est la valeur de β qui donne l'intervalle le plus court?
La longueur de l'intervalle est $\frac{X}{\beta} - \frac{X}{1-\alpha+\beta} = X\frac{10-\alpha}{\beta(1-\alpha+\beta)}$. Cet intervalle est d'autant plus petit que β est grand. Sa valeur maximale possible étant α, le plus court intervalle devient $\left[X \le \theta \le \frac{X}{\alpha}\right]$

c) Supposons que le train arrive après13 minutes d'attente:

i) Déterminer un intervalle de confiance à 90% par la formule en a) $\left[\frac{13}{1-0{,}05} \le \theta \le \frac{13}{0{,}05}\right] = [13{,}68\,;260]$

ii) Déterminer un intervalle de confiance à 90% par la formule en b) $\left[13 \le \theta \le \frac{13}{0{,}1}\right] = [13\,;130]$

iii) Déterminer un intervalle à 90% de la forme [$a \le \theta < \infty$). Est-ce que l'intervalle vous permettrait de contredire l'affirmation que θ = 10? $\left[\frac{13}{1-0{,}1} \le \theta \le \infty\right] = [14{,}44\,;\infty)$

d) Vous arrivez n fois au quai et vous notez à chaque fois votre temps d'attente. Montrer que l'intervalle suivant est un intervalle de confiance à 100(1-α) % pour θ en utilisant le fait que la durée moyenne d'attente $\bar{X}$ est approximativement de loi normale: $\left[\frac{\bar{X}}{\frac{1}{2}+\frac{z_{\alpha/2}}{\sqrt{12n}}}\ ;\ \frac{\bar{X}}{\frac{1}{2}-\frac{z_{\alpha/2}}{\sqrt{12n}}}\right]$

Si on admet que $\bar{X}$ est de loi normale, alors $\dfrac{\bar{X}-\mathrm{E}(\bar{X})}{\sqrt{\mathrm{Var}(\bar{X})}} \sim N(0;1)$. Or $\mathrm{E}(\bar{X}) = \dfrac{\theta}{2}$ et $\mathrm{Var}(\bar{X}) = \dfrac{\theta^2}{12n}$. Donc $\dfrac{\bar{X}-\mathrm{E}(\bar{X})}{\sqrt{\mathrm{Var}(\bar{X})}} =$ $\dfrac{\bar{X}-\theta/2}{\theta/\sqrt{12n}} \sim N(0;1) \Rightarrow \mathrm{P}\left[-z_{\alpha/2} \le \dfrac{\bar{X}-\theta/2}{\theta/\sqrt{12n}} \le z_{\alpha/2}\right] \approx 1 - \alpha$. Les inégalités $-z_{\alpha/2} \le \dfrac{\bar{X}-\theta/2}{\theta/\sqrt{12n}} \le z_{\alpha/2}$ peuvent être réécrites comme $\dfrac{\bar{X}}{\frac{1}{2}+\frac{z_{\alpha/2}}{\sqrt{12n}}} \le \theta \le \dfrac{\bar{X}}{\frac{1}{2}-\frac{z_{\alpha/2}}{\sqrt{12n}}}$, qui constituent alors un intervalle de confiance de niveau 1 - α.

e) Évaluer l'intervalle de confiance par la formule développée en d) à partir des 8 temps d'attente observés suivants (on admettra pour les besoins de ce numéro que n = 8 est assez « grand »):

0,05 ; 3,17 ; 4,61 ; 6,02 ; 6,44 ; 7,67 ; 9,21 ; 10,34

Intervalle de confiance : $\left[\dfrac{5{,}93875}{1/2+1{,}281552/\sqrt{12(8)}} \le \theta \le \dfrac{5{,}93875}{1/2-1{,}281552/\sqrt{12(8)}}\right] = [8{,}483;17{,}799]$

e) Dans un échantillon de taille n, l'estimateur du maximum de vraisemblance de θ est $Y = X_{(n)}$, la plus grande des n valeurs. Montrer que l'intervalle $\left[\dfrac{Y}{(1-\alpha/2)^{1/n}}\ ;\ \dfrac{Y}{(\alpha/2)^{1/n}}\right]$ est un intervalle de confiance de niveau 1-α. Utiliser cette méthode pour déterminer un intervalle de confiance à 90 % pour θ à partir des données présentées en d).

La fonction de répartition de Y est $G(y) = \dfrac{y^n}{\theta^n}$, $0 < y < \theta$, et donc $G^{-1}(u) = \theta u^{1/n}$ $0 < u < 1$. Alors

$$\mathrm{P}[G^{-1}(\alpha/2) \le Y \le G^{-1}(1-\alpha/2)] = \mathrm{P}[\theta(\alpha/2)^{1/n} < Y \le \theta(1-\alpha/2)^{1/n}] = 1-\alpha \Rightarrow \mathrm{P}\left[\frac{Y}{(1-\alpha/2)^{1/n}} \le \theta \le \frac{Y}{(\alpha/2)^{1/n}}\right] = 1-\alpha.$$

Y = 15,03659. L'intervalle de confiance est $\dfrac{15{,}03659}{(0{,}95)^{1/8}} \le \theta \le \dfrac{Y}{(0{,}05)^{1/8}}$, soit $10{,}41 \le \theta \le 15{,}04$.

5.26 Soit $X_1, \dots, X_n$ un échantillon d'une population de loi uniforme sur [0 ; θ]. On propose ici trois façons de déterminer un intervalle de confiance pour θ :

Méthode A : Soit $G(y\ ;\theta)$ la fonction de répartition de $Y = \max\{X_1, \dots, X_n\}$. L'intervalle de confiance est l'ensemble des valeurs de θ pour lesquelles $\alpha/2 \le G(y\ ;\theta) \le 1 - \alpha/2$.

Méthode B : Soit $H(z\ ;\theta)$ la fonction de répartition de $Z = \min\{X_1, \dots, X_n\}$. L'intervalle de confiance est l'ensemble des valeurs de θ pour lesquelles $\alpha/2 \le H(z\ ;\theta) \le 1 - \alpha/2$.

Méthode C : L'intervalle de confiance approximatif est l'ensemble des valeurs de θ pour lesquelles $\left|\dfrac{\bar{X}-\mu(\theta)}{\sigma(\theta)/\sqrt{n}}\right| \le z_{\alpha/2}$, où $\mu(\theta) = \theta/2$ et $\sigma(\theta) = \theta/\sqrt{12}$.

a) Méthode A

i) Montrer que $G(y) = (y/\theta)^n$, $0 \le y \le \theta$;

$$G(y) = \mathrm{P}(Y \le y) = \prod_{i=1}^{n} \mathrm{P}(X_i \le y) = \prod_{i=1}^{n}\left(\frac{y}{\theta}\right) = \left(\frac{y}{\theta}\right)^n$$

ii) Montrer que la fonction de densité de Y est $g(y) = ny^{n-1}/\theta^n$, $0 \le y \le \theta$;

$$g(y) = G'(y) = \frac{n}{\theta^n} y^{n-1},\ 0 \le y \le \theta$$

iii) Montrer que $\mathrm{E}(Y) = n\theta/(n+1)$;

$$E(Y) = \int_0^\theta y \frac{n}{\theta^n} y^{n-1} dy = \frac{n\theta}{n+1}$$

iv) Montrer que l'intervalle de confiance est $\dfrac{Y}{(1-\alpha/2)^{1/n}} \le \theta \le \dfrac{Y}{(\alpha/2)^{1/n}}$.

La probabilité que l'intervalle recouvre le paramètre est

$$\mathrm{P}\left(\frac{Y}{(1-\alpha/2)^{1/n}} \le \theta \le \frac{Y}{(\alpha/2)^{1/n}}\right) = \mathrm{P}\left(\theta(\alpha/2)^{1/n} \le Y \le \theta(1-\alpha/2)^{1/n}\right) = G[\theta(1-\alpha/2)^{1/n}] - G[\theta(\alpha/2)^{1/n}] =$$

$$\left(\frac{\theta(1-\alpha/2)^{1/n}}{\theta}\right)^n - \left(\frac{\theta(\alpha/2)^{1/n}}{\theta}\right)^n = (1-\alpha/2)-(\alpha/2) = 1\text{-}\alpha.$$

v) Soit $\mathcal{L}(Y)$ la largeur de l'intervalle de confiance. Montrer que $\mathrm{E}[\mathcal{L}(Y)]$ tend vers 0 lorsque n tend vers l'infini, que $n\mathrm{E}[\mathcal{L}(Y)]$ tend vers $\theta[\ln(1\text{-}\alpha/2) - \ln(\alpha/2)]$ et donc que $\sqrt{n}\,\mathrm{E}[\mathcal{L}(Y)]$ tend vers 0

La largeur de l'intervalle de confiance est $\mathcal{L}(Y) = Y\left[\frac{1}{(\alpha/2)^{1/n}} - \frac{1}{(1-\alpha/2)^{1/n}}\right]$ et l'espérance de la largeur est

$$\mathrm{E}[\mathcal{L}(Y)] = \frac{n\theta}{n+1}\left[\frac{1}{(\alpha/2)^{1/n}} - \frac{1}{(1-\alpha/2)^{1/n}}\right].$$

À la limite, $\lim_{n\to\infty}\mathrm{E}[\mathcal{L}(Y)] = \lim_{n\to\infty}\frac{n\theta}{n+1}\lim_{n\to\infty}\left[\frac{1}{(\alpha/2)^{1/n}} - \frac{1}{(1-\alpha/2)^{1/n}}\right] = \theta(1-1) = 0.$

$$\lim_{n\to\infty} n\mathrm{E}[\mathcal{L}(Y)] = \lim_{n\to\infty}\frac{n\theta}{n+1}\lim_{n\to\infty}\left[\frac{n}{(\alpha/2)^{1/n}} - \frac{n}{(1-\alpha/2)^{1/n}}\right] = \theta\lim_{\nu\to 0}\left[\frac{(\alpha/2)^{-\nu}-(1-\alpha/2)^{-\nu}}{\nu}\right],\text{ où } \nu = 1/n.$$

Appliquant la règle de l'hôpital,

$$\theta\lim_{\nu\to 0}\left[\frac{(\alpha/2)^{-\nu}-(1-\alpha/2)^{-\nu}}{\nu}\right] = \theta\lim_{\nu\to 0}\left[\frac{(\alpha/2)^{-\nu}[-\ln(\alpha/2)]-(1-\alpha/2)^{-\nu}[-\ln(1-\alpha/2)]}{1}\right] = \theta[\ln(1-\alpha/2)-\ln(\alpha/2)]$$

$\lim_{n\to\infty}\sqrt{n}\mathrm{E}[\mathcal{L}(Y)] = \lim_{n\to\infty}\frac{n\theta}{n+1}\lim_{n\to\infty}\left[\frac{1}{(\alpha/2)^{1/n}} - \frac{1}{(1-\alpha/2)^{1/n}}\right] = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\lim_{n\to\infty} n\mathrm{E}[\mathcal{L}(Y)] = 0\lim_{n\to\infty} n\mathrm{E}[\mathcal{L}(Y)] = 0$ puisque $\lim_{n\to\infty} n\mathrm{E}[\mathcal{L}(Y)]$ existe.

b) Méthode B

i) Montrer que $H(z) = 1-\left(1-\frac{z}{\theta}\right)^n$, $0 \le z \le \theta$.

$$H(z) = 1-\prod_{i=1}^n P(X_i > z) = 1-\prod_{i=1}^n\left(1-\frac{z}{\theta}\right) = 1-\left(1-\frac{z}{\theta}\right)^n$$

ii) Montrer que la fonction de densité de Z est $h(z) = \frac{n}{\theta}\left(1-\frac{z}{\theta}\right)^{n-1}$, $0 \le z \le \theta$.

La fonction de densité est la dérivée $h(z) = H'(z) = \frac{n}{\theta}\left(1-\frac{z}{\theta}\right)^{n-1}$, $0 \le z \le \theta$

iii) Montrer que $\mathrm{E}(Z) = \frac{\theta}{n+1}$.

$$\mathrm{E}(Z) = \int_0^\theta z\frac{n}{\theta}\left(1-\frac{z}{\theta}\right)^{n-1}dz$$

Suite à la transformation $u = z/\theta$ on obtient $\mathrm{E}(Z) = \int_0^\theta z\frac{n}{\theta}\left(1-\frac{z}{\theta}\right)^{n-1}dz = n\theta\int_0^1 u(1-u)^{n-1}du$.

Une intégrant par parties donne

$$\mathrm{E}(Z) = n\theta\int_0^1 u(1-u)^{n-1}du = \frac{\theta}{n+1}.$$

iv) Montrer que l'intervalle $\frac{Z}{1-(\alpha/2)^{1/n}} \le \theta \le \frac{Z}{1-(1-\alpha/2)^{1/n}}$ est un intervalle de confiance à 100(1-α) % pour θ.

$$\mathrm{P}\left(\frac{Z}{1-(\alpha/2)^{1/n}} \le \theta \le \frac{Z}{1-(1-\alpha/2)^{1/n}}\right) = \mathrm{P}\left([1-(1-\alpha/2)^{1/n}] \le \frac{Z}{\theta} \le [1-(\alpha/2)^{1/n}]\right) = \mathrm{P}\left(1-(1-\alpha/2)^{1/n} \le \frac{Z}{\theta} \le 1-(\alpha/2)^{1/n}\right)$$

$= \mathrm{P}\left((\alpha/2)^{1/n} \le 1 - \frac{Z}{\theta} \le (1-\alpha/2)^{1/n} \right) = \mathrm{P}\left((\alpha/2) \le \left(1 - \frac{Z}{\theta}\right)^n \le (1-\alpha/2) \right)$.

Puisque $U = H(Z) = 1 - \left(1 - \frac{Z}{\theta}\right)^n$ est de loi uniforme sur [0 ; 1], 1 - $U = \left(1 - \frac{Z}{\theta}\right)^n$ l'est aussi.

Par conséquent, $\mathrm{P}\left((\alpha/2) \le \left(1 - \frac{Z}{\theta}\right)^n \le (1-\alpha/2) \right)$ = 1 - α/2 - α/2 = 1 - α.

v) Soit $M(Z)$ la largeur de l'intervalle de confiance. Montrer que $\mathrm{E}[M(Z)]$ tend vers $\theta \frac{\ln(\alpha/2) - \ln(1-\alpha/2)}{\ln(\alpha/2)\ln(1-\alpha/2)}$ lorsque n tend vers l'infini.

La longueur de l'intervalle de confiance est M(Z) = $Z\left(\frac{1}{1-(\alpha/2)^{1/n}} - \frac{1}{1-(1-\alpha/2)^{1/n}} \right)$ et son espérance est

$\mathrm{E}(M(Z)) = \frac{\theta}{n+1}\left(\frac{1}{1-(1-\alpha/2)^{1/n}} - \frac{1}{1-(\alpha/2)^{1/n}} \right)$.

Alors $\lim_{n\to\infty} \mathrm{E}(M(Z)) = \lim_{n\to\infty} \frac{\theta}{n+1}\left(\frac{1}{1-(1-\alpha/2)^{1/n}} - \frac{1}{1-(\alpha/2)^{1/n}} \right)$. La limite est $\theta \frac{\ln(1-\alpha/2) - \ln(\alpha/2)}{\ln(1-\alpha/2)\ln(\alpha/2)}$.

c) Méthode C

i) Montrer que l'intervalle de confiance approximatif est $\frac{2\bar{X}}{1 + \frac{2z_{\alpha/2}}{\sqrt{12n}}} \le \theta \le \frac{2\bar{X}}{1 - \frac{2z_{\alpha/2}}{\sqrt{12n}}}$.

$$\mathrm{P}\left(\frac{2\bar{X}}{1 + \frac{2z_{\alpha/2}}{\sqrt{12n}}} \le \theta \le \frac{2\bar{X}}{1 - \frac{2z_{\alpha/2}}{\sqrt{12n}}} \right) = \mathrm{P}\left(\frac{\theta}{2}\left(1 - \frac{2z_{\alpha/2}}{\sqrt{12n}}\right) \le \bar{X} \le \frac{\theta}{2}\left(1 - \frac{2z_{\alpha/2}}{\sqrt{12n}}\right) \right) = \mathrm{P}\left(-\frac{\theta z_{\alpha/2}}{\sqrt{12n}} \le \bar{X} - \frac{\theta}{2} \le \frac{\theta z_{\alpha/2}}{\sqrt{12n}} \right)$$

$$= \mathrm{P}\left(-\frac{\theta z_{\alpha/2}}{\sqrt{12n}} \le \frac{\bar{X} - \theta/2}{} \le \frac{\theta z_{\alpha/2}}{\sqrt{12n}} \right) = \mathrm{P}\left(-z_{\alpha/2} \le \frac{\bar{X} - \theta/2}{\theta/\sqrt{12n}} \le z_{\alpha/2} \right) = 1 - \alpha.$$

Si on suppose que $\bar{X}$ est de loi normale, alors

$\left(-z_{\alpha/2} \le \frac{\bar{X} - \theta/2}{\theta/\sqrt{12n}} \le z_{\alpha/2} \right)$ est de loi normale centrée réduite et donc $\mathrm{P}\left(-z_{\alpha/2} \le \frac{\bar{X} - \theta/2}{\theta/\sqrt{12n}} \le z_{\alpha/2} \right) = 1 - \alpha$.

ii) Montrer que l'espérance de la largeur $K(\bar{X})$ de l'intervalle de confiance est $\frac{\theta z_{\alpha/2}}{\sqrt{3n}/2 - z_{\alpha/2}^2/\sqrt{12n}}$.

La largeur $K(\bar{X})$ est égale à $\frac{2\bar{X}}{1 - \frac{2z_{\alpha/2}}{\sqrt{12n}}} - \frac{2\bar{X}}{1 - \frac{2z_{\alpha/2}}{\sqrt{12n}}}$ et l'espérance de la largeur est

$$\mathrm{E}[K(\bar{X})] = \theta\left(\frac{1}{1 - 2z_{\alpha/2}/\sqrt{12n}} - \frac{1}{1 + 2z_{\alpha/2}/\sqrt{12n}} \right) = \theta \frac{4z_{\alpha/2}/\sqrt{12n}}{1 - 2z_{\alpha/2}/\sqrt{12n}} = \theta \frac{4z_{\alpha/2}/\sqrt{12n}}{1 - 4z_{\alpha/2}^2/12n} = \theta \frac{4z_{\alpha/2}/\sqrt{n}}{\sqrt{12} - 4z_{\alpha/2}^2/\sqrt{12n}}.$$

iii) Montrer que $\mathrm{E}[K(\bar{X})]$ tend vers 0 lorsque n tend vers l'infini et que $\sqrt{n}\mathrm{E}[K(\bar{X})]$ tend vers $\frac{2\theta z_{\alpha/2}}{\sqrt{3}}$.

$$\lim_{n\to\infty} \mathrm{E}[K(\bar{X})] = \lim_{n\to\infty} \theta \frac{4z_{\alpha/2}/\sqrt{n}}{\sqrt{12} - 4z_{\alpha/2}^2/\sqrt{12n}} = 0 \; ; \; \lim_{n\to\infty} \mathrm{E}[\sqrt{n}K(\bar{X})] = \lim_{n\to\infty} \theta \frac{4z_{\alpha/2}}{\sqrt{12} - 4z_{\alpha/2}^2/\sqrt{12n}} = \theta \frac{4z_{\alpha/2}}{\sqrt{12}} = \frac{2\theta z_{\alpha/2}}{\sqrt{3}}.$$

d) L'une des trois méthodes est inacceptable. Laquelle? Des deux autres méthodes, laquelle est meilleure?
La méthode B est inadmissible. On s'attend normalement à ce que l'intervalle se rétrécisse à mesure que n augmente et tende vers 0 lorsque n tend vers l'infini.
La méthode C est acceptable, car elle satisfait à cette condition.
La méthode A est encore meilleure, car non seulement la largeur tend-t-elle vers 0 lorsque n tend vers l'infini, mais la largeur, multipliée par $\sqrt{n}$ tend elle aussi vers 0.

e) On conçoit aisément pourquoi la méthode B est si peu efficace, comparée à la méthode A: le minimum ne peut contenir beaucoup d'information sur θ, qui, elle, est la valeur maximale de la variable.

i) Vérifier que $\hat{\theta}_1 = (n+1)Z$ est sans biais et montrer que $\mathrm{Var}(\hat{\theta}_1) = \dfrac{n\theta^2}{n+2}$ [Remarquer que la variance ne tend pas vers 0 lorsque $n\to\infty$].

$$\mathrm{E}(Z) = \frac{\theta}{n+1} \Rightarrow \mathrm{E}(\hat{\theta}_1) = \mathrm{E}[(n+1)Z] = (n+1)\mathrm{E}[Z] = (n+1)\frac{\theta}{n+1} \Rightarrow \theta\ .$$

$$\mathrm{E}(Z^2) = \int_0^\theta \frac{n}{\theta}z^2\left(1-\frac{z}{\theta}\right)^{n-1}dz = n\theta\int_0^\theta\left(\frac{z^2}{\theta^2}\right)\left(1-\frac{z}{\theta}\right)^{n-1}dz = n\theta^2\int_0^1 u^2(1-u)^{n-1}du = \frac{2\theta^2}{(n+2)(n+1)}\ .$$

$$\mathrm{Var}(Z) = E(Z^2) - [E(Z)]^2 = \frac{2\theta^2}{(n+2)(n+1)} - \left(\frac{\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+2)(n+1)^2}\ .$$

Donc $\mathrm{Var}(\hat{\theta}_1) = \mathrm{Var}[(n+1)Z] = (n+1)^2\mathrm{Var}[Z] = (n+1)^2\dfrac{n\theta^2}{(n+2)(n+1)^2} = \dfrac{n\theta^2}{(n+2)}$.

ii) Vérifier que $\hat{\theta}_2 = \dfrac{n+1}{n}Y$ est sans biais et montrer que $\mathrm{Var}(\hat{\theta}_2) = \dfrac{\theta^2}{n(n+2)}$.

$$\mathrm{E}(\hat{\theta}_2) = \mathrm{E}\left(\frac{n+1}{n}Y\right) = \frac{n+1}{n}\mathrm{E}(Y) = \frac{n+1}{n}\frac{n\theta}{n+1} = \theta\ .$$

$$\mathrm{E}(Y^2) = \int_0^\theta y\frac{ny^{n-1}}{\theta^n} = \frac{n\theta^2}{n+2}\ . \quad \mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{n\theta^2}{n+2} - \left(\frac{n\theta}{n+1}\right)^2 = \frac{n\theta^2}{(n+1)^2(n+2)}\ .$$

$$\mathrm{Var}(\hat{\theta}_2) = \frac{(n+1)^2}{n^2}\mathrm{Var}(Y) = \frac{(n+1)^2}{n^2}\frac{n\theta^2}{(n+1)(n+2)} = \frac{\theta^2}{n(n+2)}\ .$$

iii) Vérifier que $\hat{\theta}_3 = 2\bar{X}$ est sans biais et montrer que $\mathrm{Var}(\hat{\theta}_3) = \dfrac{\theta^2}{3n}$.

$\mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$, où $\sigma^2 = \theta^2/12$ et $\mathrm{Var}(\bar{X}) = \dfrac{\theta^2}{12n}$ et $\mathrm{Var}(\hat{\theta}_3) = \mathrm{Var}(2\bar{X}) = \dfrac{4\theta^2}{12n} = \dfrac{\theta^2}{3n}$.

f) Un échantillon de taille $n = 30$ est tiré d'une population de loi uniforme sur [0 ; θ]. Voici les observations:

0,15	3,82	8,31	12,18	19,61	21,07	27,89	31,36	36,53	42,21
0,22	4,23	8,46	15,33	20,70	21,96	29,46	31,43	38,68	48,64
2,52	4,60	10,22	18,74	20,83	26,00	31,33	36,38	41,98	49,39

Déterminer un intervalle de confiance utilisant chacune des trois méthodes.

Données de base: y = 49, 390; z = 0,150; $\bar{x}$ = 22, 141.

Intervalles de confiance à 95 %:

Méthode A: [55,85 ; 49,43]

Méthode B: [1,30 ; 177,82]

Méthode C: [55,81 ; 36,70]