T. D. n$^{\text{o}}$ 1
SAS à probabilités égales

Exercice 1. Des raisonnements classiques dans la théorie des sondages.

L'exercice propose de démontrer des résultats présentés dans le cours et d'insister sur des techniques de raisonnement usuelles en sondage.

Considérons que nous voulions estimer le total et la moyenne d'une grandeur Y dans une population U de taille N. Pour cela, nous procédons à un **sondage aléatoire simple sans remise** de taille n et on note $\mathcal{S}$ l'échantillon aléatoire obtenu.

1. Combien y a-t-il d'échantillons possibles ? Quelle est la probabilité de tirer chacun d'entre eux ?
2. Nous considérons un individu k quelconque dans U. Combien y a-t-il d'échantillons contenant cet individu ? En déduire la probabilité de tirage de k.
3. Nous rappelons que nous notons δ_i la variable aléatoire valant 1 si i appartient à l'échantillon et 0 sinon.
   (i) Que vaut $\mathbb{E}\left(\delta_i\right)$?
   (ii) Comment pouvons-nous réécrire $\sum_{i \in \mathcal{S}} Y_i$ à partir des δ_i ?
4. En déduire que :
   (i) $\widehat{T}_n = \dfrac{N}{n} \sum_{i \in \mathcal{S}} Y_i$ estime sans biais le total T_Y,
   (ii) et que $\widehat{\mu}_n = \dfrac{1}{n} \sum_{i \in \mathcal{S}} Y_i$ estime sans biais la moyenne μ_Y.
5. Combien y a-t-il d'échantillons comprenant les individus identifiés k et l ? En déduire la probabilité de tirer ces deux individus conjointement. Que vaut alors $\mathbb{E}\left(\delta_k \delta_l\right)$? En déduire $\mathrm{Cov}\left(\delta_k \delta_l\right)$.
6. Nous rappelons que nous notons $\sigma^2_{Y,c} = \dfrac{1}{N-1} \sum_{k=1}^{N} \left(Y_k - \mu_Y\right)^2$ et $f = \dfrac{n}{N}$.

   Montrer que
   (i)
   $$\mathrm{Var}\left(\widehat{T}_n\right) = N^2 (1-f) \frac{\sigma_c^2}{n}$$
   (ii) et que
   $$\mathrm{Var}\left(\,\widehat{\mu}_n\,\right) = (1-f)\,\frac{\sigma_c^2}{n}\cdot$$
7. Quel est l'intérêt du sondage sans remise par rapport au sondage avec remise ?

Exercice 2. Des formules du cours.

Démontrer, lors d'un tirage avec remise, que :

— $\mathbb{E}\left(S^2_{n,c}\right) = \sigma^2$ c'est-à-dire que $S^2_{n,c}$ est un estimateur sans biais de σ^2 lors d'un tirage avec remise ;

— $\mathrm{Var}\left(S^2_{n,c}\right) = \dfrac{1}{n(n-1)}\left[(n-1)\mu_4 - (n-3)\sigma^4\right]$.

Démontrer, lors d'un tirage sans remise, que :

$$\begin{aligned}\mathrm{Var}\left(\widehat{\sigma}^2_n\right) &= \frac{(N-n)}{n(n-1)N(N-2)(N-3)}\Big(\mu_4(N-1)\left[N(n-1)-(n+1)\right] \\ &\quad -\sigma^4\left[N^2(n-3)+6N-3(n+1)\right]\Big).\end{aligned}$$

Exercice 3. Échantillonnage sans remise et avec remise.

Soit une variable Y définie sur une population de taille $N = 4$ individus.

i	1	2	3	4
Y	11	10	8	11

1. Quelle est la distribution de fréquence de Y ? En déduire $\mu_Y = \mathbb{E}(Y)$, $\sigma^2_Y = \mathrm{Var}(Y)$, et $\sigma^2_{Y,c}$.
2. Nous tirons un échantillon de taille $n =$ **2 sans remise**, à probabilités égales.
   (i) Combien d'échantillons pouvons-nous tirer ?
   (ii) Pour chaque échantillon, calculer sa moyenne $\widehat{\mu}_n$ et $S^2_{n,c}$ sa variance corrigée.
   (iii) Quelle est la distribution de $\widehat{\mu}_n$ et de $S^2_{n,c}$? En déduire $\mathbb{E}\left(\widehat{\mu}_n\right)$, $\mathrm{Var}\left(\widehat{\mu}_n\right)$, $\mathbb{E}\left(S^2_{n,c}\right)$ et $\mathrm{Var}\left(S^2_{n,c}\right)$.
   (iv) Commenter, c'est-à-dire comparer et analyser les résultats trouvés.
3. Nous tirons un échantillon de taille $n =$ **2 avec remise**, à probabilités égales. Répondre aux mêmes questions qu'à la question 2.
4. Nous tirons un échantillon de taille $n =$ **1** à probabilités égales. Répondre aux mêmes questions qu'à la question 2.

Exercice 4. Application immédiate des formules du cours.

Dans un groupe de 80 étudiants nous tirons au hasard à probabilités égales sans remise un échantillon de taille n. Nous prendrons $n = 4$ puis $n = 40$.

1. Nous observons dans l'échantillon la variable aléatoire Y, dépense hebdomadaire pour la culture. Nous trouvons $\widehat{\mu}_n(obs) = 12$ euros et $S(obs) = 6$ euros. Donner dans les deux cas, une estimation de la dépense moyenne dans le groupe et la précision de cette estimation.
2. Nous observons dans l'échantillon 75% de femmes. Donner dans les deux cas, une estimation de la proportion de femmes dans le groupe et la précision de cette estimation.
3. Commenter les résultats obtenus.

Exercice 5. Vérification dans un cas simple.

Une entreprise possède cinq succursales. Un inspecteur ne peut en examiner que deux par tournées. Dans chaque succursale, nous mesurons une variable d'intérêt Y (Nombre de nouveaux clients dans l'année). La situation réelle des cinq succursales est la suivante :

$$Y_1 = 100; \quad Y_2 = 80; \quad Y_3 = 100; \quad Y_4 = 120; \quad Y_5 = 90.$$

1. énumérer tous les échantillons possibles correspondant à une tournée et pour chaque échantillon calculer $\widehat{\mu}_n$, l'estimateur de la moyenne μ_Y ainsi obtenu.
2. Vérifier que $\widehat{\mu}_Y$ est un estimateur sans biais de la moyenne μ_Y.
3. Calculer la variance de $\widehat{\mu}_Y$, directement et en utilisant les formules du cours.
4. Pour chaque échantillon possible, calculer sa variance corrigée $S^2_{n,c}$. Vérifier que $S^2_{n,c}$ est un estimateur sans biais de la variance corrigée σ^2_c de la population.

Exercice 6. D'après le livre « Exercices de sondage » de A.-M. Dussaix et J.-M. Grosbras.

200 ménages touristiques séjournant en Bretagne ont dépensé, en moyenne journalière, 130 euros. De plus, l'écart-type estimé s'élève à 35 euros. Sachant que dans la Bretagne il est venu 10 000 ménages touristiques, que pouvons-nous dire de la dépense totale journalière de l'ensemble de ces ménages ? Nous supposons que l'échantillon est issu d'un sondage aléatoire à probabilités égales sans remise.

Exercice 7. Extrait du livre « Exercices corrigés de méthodes de sondage » de P. Ardilly et Y. Tillé.

Nous voulons estimer la surface moyenne cultivée dans les fermes d'un canton rural. Sur $N = 2010$ fermes que comprend ce canton, nous en tirons 100 par sondage aléatoire simple. Nous mesurons y_i la surface cultivée dans la ferme i en hectares, et nous trouvons :

$$\sum_{i \in S} y_i = 2907\,\mathrm{ha}$$

et

$$\sum_{i \in S} y_i^2 = 154593\,\mathrm{ha}^2.$$

1. Donner la valeur de l'estimateur sans biais classique de la moyenne de la population.
2. Donner un intervalle de confiance à 95% pour la moyenne. Nous rappelons que le quantile d'une loi normale à 95% est approximativement égal à $1,96$.

Exercice 8. Probabilité d'inclusion.

Soit la population $\{1, 2, 3\}$ et le plan de sondage suivant :

$$\mathbb{P}(\{1,2\}) = \frac{1}{2} \quad \mathbb{P}(\{1,3\}) = \frac{1}{4} \quad \mathbb{P}(\{2,3\}) = \frac{1}{4}.$$

1. Est-ce un sondage aléatoire simple ?
2. Calculer π_1, π_2 et π_3 les probabilités d'inclusion d'ordre 1.
3. Calculer π_{12} et π_{23} les probabilités d'inclusion d'ordre 2.
4. Quel est le π-estimateur de $\overline{Y}$? Nous noterons Y_1, Y_2 et Y_3 les valeurs respectives de la variable Y.
   a) si l'échantillon $\{1, 2\}$ est tiré ?
   b) si l'échantillon $\{1, 3\}$ est tiré ?
   c) si l'échantillon $\{2, 3\}$ est tiré ?
5. Vérifier que le π-estimateur est un estimateur sans biais.
6. Écrire ce que seraient les probabilités d'échantillon $\mathbb{P}$ et les probabilités d'inclusion π pour un sondage aléatoire simple à probabilités égales sans remise.