Université du Québec à Montréal

Devoir I

Travail présenté à
Michaël Lalancette

Dans le cadre du cours
Statistique informatique
STT3010

Par
Galiane Charbonneau
CHAG68010004

Le 1 février 2023

Question 1

Soit $X \sim p$ une variable aléatoire discrète. On note $\{x_1, \ldots, x_k\} \subset \mathbb{R}$ le support de sa loi et $p_i = p(x_i) = \mathbb{P}(X = x_i)$ pour $1 \leqslant i \leqslant k$. Afin de pouvoir générer à partir de la loi de X, définissons la variable aléatoire $U \sim \mathscr{U}(0,1)$ et la fonction

$$\begin{aligned} h\colon [0,1] &\longrightarrow \mathbb{R} \\ u &\mapsto \begin{cases} x_1, & u \leqslant p_1 \\ x_2, & p_1 < u \leqslant p_1 + p_2 \\ \vdots \\ x_{k-1}, & \sum_{j=1}^{k-2} p_j < u \leqslant \sum_{j=1}^{k-1} p_j \\ x_k, & \sum_{j=1}^{k-1} p_j < u \leqslant 1 \end{cases} \end{aligned}$$

En classe, nous avons démontré que $h(U) \sim g$. On définit maintenant la variable aléatoire

$$L = \mathbb{1}(U \leqslant p_1) + \sum_{j=2}^{k} j \mathbb{1}\left(\sum_{i=1}^{j-1} p_i < U \leqslant \sum_{i=1}^{j} p_i\right)$$

Cette variable aléatoire compte le nombre d'itérations requises afin de pouvoir simuler une observation de la loi g à l'aide de l'algorithme présenté à la page 7 du chapitre 2 des notes de cours.

a) Calculons $\mathbb{E}[L]$. On a

$$\begin{aligned} \mathbb{E}[L] &= \mathbb{E}\left[\mathbb{1}(U \leqslant p_1) + \sum_{j=2}^{k} j \mathbb{1}\left(\sum_{i=1}^{j-1} p_i < U \leqslant \sum_{i=1}^{j} p_i\right)\right] \\ &= \mathbb{E}[\mathbb{1}(U \leqslant p_1)] + \sum_{j=2}^{k} j \mathbb{E}\left[\mathbb{1}\left(\sum_{i=1}^{j-1} p_i < U \leqslant \sum_{i=1}^{j} p_i\right)\right], \quad \text{par linéarité de l'espérance} \\ &= \mathbb{P}(U \leqslant p_1) + \sum_{j=2}^{k} j \mathbb{P}\left(\sum_{i=1}^{j-1} p_i < U \leqslant \sum_{i=1}^{j} p_i\right) \\ &= p_1 + \sum_{j=2}^{k} j \left(\sum_{i=1}^{j} p_i - \sum_{i=1}^{j-1} p_i\right), \quad \text{car } U \sim \mathscr{U}(0,1) \\ &= \sum_{j=1}^{k} j p_j \end{aligned}$$

b) Démontrons maintenant que l'espérance calculée en a) est minimisée si $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_k$. Procédons en plusieurs étapes.

(i) Fixons d'abord $\sigma\colon \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$ une application bijective (aussi appelée permutation).

(ii) Introduisons une notation qui sera utile plus tard. On définit

$$\begin{aligned} \mu_1 &= \sigma^{-1}(1) \\ \mu_j &= \mu_{j-1} - (j-1) + \sigma^{-1}(j), \quad j \in \{2, \ldots, k\} \end{aligned}$$

où σ^{-1} représente l'application réciproque de σ (qui existe bel et bien, car σ est une bijection).

(iii) Définissons ces mêmes quantités, mais d'une manière alternative. On définit

$$\begin{aligned} \gamma_1 &= \sigma^{-1}(1) \\ \gamma_j &= \sigma^{-1}(j) + \sum_{i=1}^{j-1} (\sigma^{-1}(i) - i), \quad j \in \{2, \ldots, k\} \end{aligned} \tag{1}$$

(iv) Démontrons maintenant que les définitions (ii) et (iii) sont équivalentes. D'abord, on voit immédiatement que $\mu_1 = \gamma_1$. Ensuite, pour $j = 2$, on a que

$$\begin{aligned}
\mu_2 &= \mu_1 - 1 + \sigma^{-1}(2) \\
&= \sigma^{-1}(1) - 1 + \sigma^{-1}(2) \\
&= \sigma^{-1}(2) + \sum_{i=1}^{2-1} (\sigma^{-1}(i) - i) \\
&= \gamma_2
\end{aligned}$$

Supposons maintenant que pour un certain $j < k$, on ait $\gamma_j = \mu_j$. On a alors que

$$\begin{aligned}
\mu_{j+1} &= \mu_j - j + \sigma^{-1}(j+1) \\
&= \gamma_j - j + \sigma^{-1}(j+1), \quad \text{par hypothèse d'induction} \\
&= \sigma^{-1}(j) + \sum_{i=1}^{j-1} (\sigma^{-1}(i) - i) - j + \sigma^{-1}(j+1) \\
&= \sigma^{-1}(j+1) + \sum_{i=1}^{j} (\sigma^{-1}(i) - i) \\
&= \gamma_{j+1}
\end{aligned}$$

Ainsi, par le principe d'induction, on a bien que $\mu_j = \gamma_j \ \forall j \in \{1, \ldots, k\}$.

(v) Démontrons que pour tout $j \in \{1, \ldots, k\}$, $\gamma_j \geqslant j$. (Dans cette section, il est plus facile de travailler avec la définition (iii). C'est ce qui a motivé celle-ci). D'abord, il est évident que $\gamma_1 = \sigma^{-1}(1) \geqslant 1$ (en effet, $\sigma^{-1}(1) \in \{1, \ldots, k\}$). Pour la suite, procédons par contradiction. Supposons qu'il existe $j \in \{2, \ldots, k\}$ tel que $\gamma_j < j$.

On a alors la suite d'équivalences suivante:

$$\begin{aligned}
\gamma_j < j &\iff \sigma^{-1}(j) + \sum_{i=1}^{j-1}(\sigma^{-1}(i) - i) < j \\
&\iff \sum_{i=1}^{j} \sigma^{-1}(i) - \sum_{i=1}^{j-1} i < j \\
&\iff \sum_{i=1}^{j} \sigma^{-1}(i) - \frac{j(j-1)}{2} < j \\
&\iff \sum_{i=1}^{j} \sigma^{-1}(i) < j + \frac{j(j-1)}{2} \\
&\iff \sum_{i=1}^{j} \sigma^{-1}(i) < \frac{j(j+1)}{2} \\
&\iff \sum_{i=1}^{j} \sigma^{-1}(i) < \sum_{i=1}^{j} i
\end{aligned}$$

Cette dernière inégalité est la contradiction recherchée. En effet, le terme de gauche représente la somme de $2 \leqslant j \leqslant k$ entiers choisis dans l'ensemble $\{1, \ldots, k\}$. Elle doit donc nécessairement être au moins égale à la somme des j plus petits entiers de cet ensemble, à savoir les entiers $\{1, 2, \ldots, j\}$.

(vi) Voici quelques dernières observations. D'abord,

$$\begin{aligned}
\mu_k &= \gamma_k \\
&= \sigma^{-1}(k) + \sum_{i=1}^{k-1}(\sigma^{-1}(i) - i) \\
&= \sum_{i=1}^{k} \sigma^{-1}(i) - \sum_{i=1}^{k-1} i \\
&= \sum_{i=1}^{k} i - \sum_{i=1}^{k-1} i \quad \text{car } \sigma \text{ est une bijection} \\
&= k
\end{aligned} \tag{2}$$

Ensuite, pour tout $j \in \{1, \ldots, k-1\}$

$$\begin{aligned}
\mu_j p_j + \sigma^{-1}(j+1)p_{j+1} &= (\mu_j - j + j)p_j + \sigma^{-1}(j+1)p_{j+1} \\
&= (\mu_j - j)p_j + jp_j + \sigma^{-1}(j+1)p_{j+1} \\
&\geqslant jp_j + \big[(\mu_j - j) + \sigma^{-1}(j+1)\big]p_{j+1} \quad \text{car } p_j \geqslant p_{j+1} \text{ et } \mu_j \geqslant j \\
&= jp_j + \mu_{j+1}p_{j+1}
\end{aligned} \tag{3}$$

(vii) Nous sommes maintenant en mesure de démontrer le résultat d'intérêt. On a

$$\begin{aligned}
\sum_{j=1}^{k} j p_{\sigma(j)} &= \sum_{j=1}^{k} \sigma^{-1}(j) p_j \\
&= \mu_1 p_1 + \sigma^{-1}(2) p_2 + \sum_{j=3}^{k} \sigma^{-1}(j) p_j \\
&\geqslant p_1 + \mu_2 p_2 + \sum_{j=3}^{k} \sigma^{-1}(j) p_j, \quad \text{par (3)} \\
&\geqslant p_1 + 2p_2 + \mu_3 p_3 + \sum_{j=4}^{k} \sigma^{-1}(j) p_j, \quad \text{par (3)} \\
&\vdots \\
&\geqslant \sum_{j=1}^{k-2} j p_j + \mu_{k-1} p_{k-1} + \sigma^{-1}(k) p_k, \quad \text{par (3)} \\
&\geqslant \sum_{j=1}^{k-1} j p_j + \mu_k p_k, \quad \text{par (3)} \\
&= \sum_{j=1}^{k} j p_j, \quad \text{par (2)}
\end{aligned}$$

Comme la permutation σ était arbitraire, on a bien que

$$\underset{\sigma \in S_k}{\arg\min} \sum_{j=1}^{k} j p_{\sigma(j)} = \sigma_{\text{id}}$$

où S_k représente l'ensemble des $k!$ permutations de $\{1, \ldots, k\}$ et $\sigma_{\text{id}} \in S_k$ est l'application identité.

c) Les figures 1 et 2 comparent les distributions des temps requis par l'ordinateur pour simuler $m = 10$ échantillons de taille $n = 5000$ de la loi Binomiale$(1000, 0.9)$ selon la façon dont sont ordonnées les probabilités en entrée de l'algorithme.

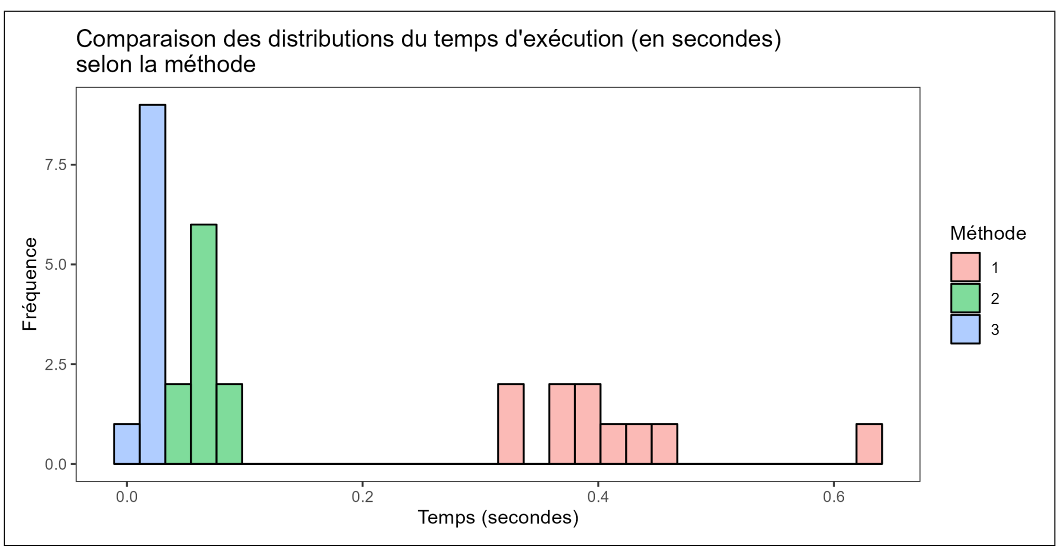

Figure 1:

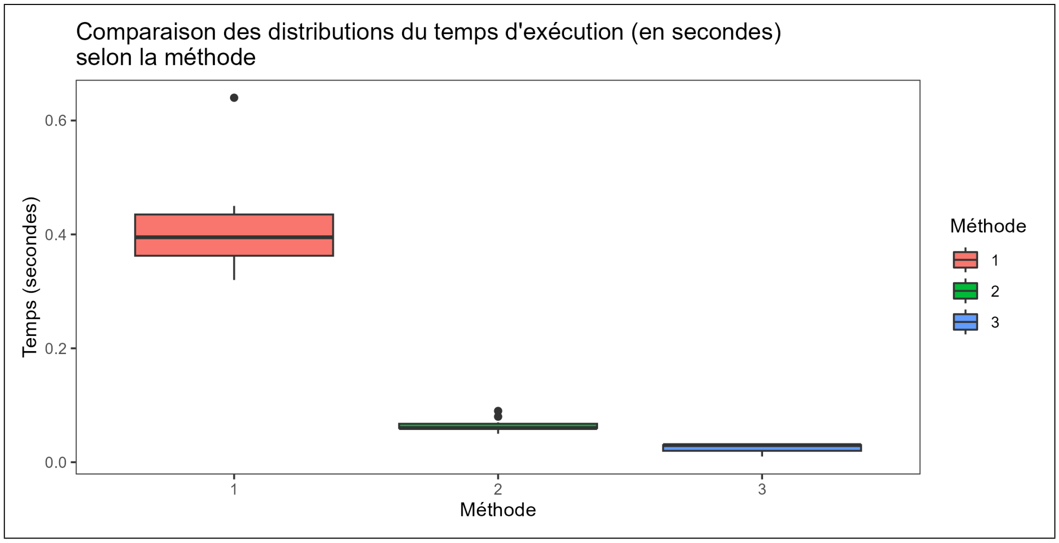

Figure 2:

Méthode	Moyenne	Écart-type
1	0.412	0.091
2	0.064	0.0126
3	0.025	0.00707

Table 1: Statistiques descriptives sur les temps d'exécution

A priori, le temps d'exécution dépend fortement de la façon dont les probabilités sont ordonnées en entrée de l'algorithme. Afin de tester si les différences de médianes sont statistiquement significatives, effectuons un test de Kruskal Wallis (pendant non paramétrique d'une ANOVA). La valeur-p réalisée étant de 2.244×10^{-6}, on peut rejeter avec confiance l'hypothèse selon laquelle la médiane du temps d'exécution est la même quelle que soit la méthode utilisée (ou, de manière équivalente, la façon dont les probabilités sont ordonnées au début de l'algorithme). Plus spécifiquement, remarquons que les méthodes 2 et 3 donnent lieu à des performances très similaires. Cela s'explique par le fait que la loi binomiale qui a été simulée possède un paramètre $p = 0.9$ proche de 1. Étant donné que les réalisations les plus probables se situent autour de 900, le fait d'ordonner les valeurs du support de la loi de manière décroissante implique que ces valeurs les plus probables font partie des premières à être testées par l'algorithme. Néanmoins, le temps moyen requis en utilisant cette méthode est significativement plus élevé que celui de la méthode 3. Cette affirmation découle de l'observation que la réalisation maximale associée à la troisième méthode est inférieure à la réalisation minimale associée à la deuxième méthode. Ainsi, si l'on effectuait, par exemple, un test de permutation (non paramétrique) pour tester l'égalité des moyennes (des méthodes 2 et 3), la valeur de la statistique de test observée (différence des moyennes) prendrait sa valeur la plus extrême. Ces résultats sont tout à fait cohérents avec le résultat démontré plus tôt.

Question 2

a) Identifions tout d'abord la loi marginale de la variable aléatoire A. Il est clair que A suit une loi de Bernouilli. Il suffit donc de calculer son espérance pour caractériser sa loi. On a

$$\begin{aligned}
\mathbb{P}(A=1) &= \int_{\mathbb{R}} \mathbb{P}(A=1 \mid Y=y) g(y)\, dy \\
&= \int_{\mathbb{R}} \frac{f(y)}{Mg(y)} g(y)\, dy \\
&= \frac{1}{M} \int_{\mathbb{R}} f(y)\, dy \\
&= \frac{1}{M}
\end{aligned}$$

$\therefore A \sim \text{Bern}\left(\frac{1}{M}\right)$

Ainsi,

$$\begin{aligned}
\mathbb{E}[h(Y)] &= \mathbb{E}\big[\mathbb{E}[h(Y) \mid A]\big], \quad \text{par le théorème de l'espérance totale} \\
&= \mathbb{E}[h(Y) \mid A=1]\mathbb{P}(A=1) + \mathbb{E}[h(Y) \mid A=0]\mathbb{P}(A=0) \\
&= \mathbb{E}[h(X)]\frac{1}{M} + \mathbb{E}[h(Y) \mid A=0]\left(1 - \frac{1}{M}\right) \\
&= \frac{\theta}{M} + \mathbb{E}[h(Y) \mid A=0]\left(1 - \frac{1}{M}\right)
\end{aligned}$$

Il suffit d'isoler $\mathbb{E}[h(Y) \mid A=0]$ et on obtient

$$\mathbb{E}[h(Y) \mid A=0] = \left(1 - \frac{1}{M}\right)^{-1} \left\{\mathbb{E}[h(Y)] - \frac{\theta}{M}\right\}$$

ce qui permet d'écrire

$$\theta = M\mathbb{E}[h(Y)] - (M-1)\mathbb{E}[h(Y) \mid A=0]$$

b) Commençons par identifier la loi du vecteur $\mathbf{A} = (A_1, \ldots, A_n)$. Comme $A_i \sim \mathrm{Bern}\big(\frac{1}{M}\big)$ $\forall i \in \{1, \ldots, n\}$ et qu'elles sont indépendantes, on a que

$$\begin{aligned}
\mathbb{P}(A_1 = a_1, \ldots, A_n = a_n) &= \prod_{i=1}^{n} \mathbb{P}(A_i = a_i) \\
&= \prod_{i=1}^{n} \left(\frac{1}{M}\right)^{a_i} \left(1 - \frac{1}{M}\right)^{1-a_i} \mathbb{1}(a_i \in \{0,1\}) \\
&= \left(\frac{1}{M}\right)^{\sum_{i=1}^n a_i} \left(1 - \frac{1}{M}\right)^{n-\sum_{i=1}^n a_i} \mathbb{1}\big((a_1, \ldots, a_n) \in \{0,1\}^n\big)
\end{aligned}$$

Calculons maintenant $\mathbb{E}\big[\hat{\theta}_a \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\big]$ et $\mathbb{V}\mathrm{ar}\big[\hat{\theta}_a \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\big]$. On a

$$\begin{aligned}
\mathbb{E}\left[\hat{\theta}_a \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\right] &= \mathbb{E}\left[\mathbb{E}[\hat{\theta}_a \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\right] \\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i = 1)} \sum_{i=1}^n \mathbb{E}[h(Y_i) \mid \mathbf{A}] \mathbb{1}(A_i = 1) \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\right] \\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i = 1)} \sum_{i=1}^n \mathbb{E}[h(Y_i) \mid A_i] \mathbb{1}(A_i = 1) \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\right] \\
&= \sum_{\substack{(a_1, \ldots, a_n) \in \{0,1\}^n \\ \sum_{i=1}^n a_i > 0}} \left\{ \frac{1}{\sum_{i=1}^n \mathbb{1}(a_i = 1)} \sum_{i=1}^n \mathbb{E}[h(Y_i) \mid A_i = a_i] \mathbb{1}(a_i = 1) \right. \\
&\qquad\qquad \left. \times \mathbb{P}\left(\mathbf{A} = (a_1, \ldots, a_n) \mid \sum_{i=1}^n a_i > 0\right) \right\} \\
&= \sum_{k=1}^n \binom{n}{k} \frac{1}{k} \times k \mathbb{E}[h(Y) \mid A = 1] \times \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{n-k} \left[1 - \left(1 - \frac{1}{M}\right)^n\right]^{-1} \\
&= \mathbb{E}[h(Y) \mid A = 1] \left[1 - \left(1 - \frac{1}{M}\right)^n\right]^{-1} \sum_{k=1}^n \binom{n}{k} \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{n-k} \\
&= \mathbb{E}[h(Y) \mid A = 1] \left[1 - \left(1 - \frac{1}{M}\right)^n\right]^{-1} \left[1 - \left(1 - \frac{1}{M}\right)^n\right] \\
&= \mathbb{E}[h(Y) \mid A = 1] \\
&= \mathbb{E}[h(X)] \\
&= \theta
\end{aligned}$$

$\therefore$ Sachant l'événement $\big\{\sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\big\}$, l'estimateur $\hat{\theta}_a$ est sans biais pour θ.

Pour la variance, remarquons d'abord que $\forall \mathbf{a} = (a_1, \ldots, a_n) \in \{0,1\}^n \setminus \{(0, \ldots, 0)\}$,

$$\mathbb{E}[\hat{\theta}_a \mid \mathbf{A} = \mathbf{a}] = \frac{1}{k} \times k\mathbb{E}[h(Y) \mid A = 1] = \mathbb{E}[h(Y) \mid A = 1]$$

où $k = \sum_{i=1}^n \mathbb{1}(a_i = 1) > 0$. Ainsi,

$$\mathbb{P}\left(\mathbb{E}[\hat{\theta}_a \mid \mathbf{A}] = \mathbb{E}[h(Y) \mid A = 1] \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\right) = 1$$

Par conséquent,

$$\mathbb{V}\mathrm{ar}\left[\mathbb{E}[\hat{\theta}_a \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\right] = 0$$

On a donc que

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\hat{\theta}_a \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right] &= \mathbb{E}\left[\mathbb{V}\mathrm{ar}[\hat{\theta}_a \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}[\hat{\theta}_a \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right] \\
&= \mathbb{E}\left[\mathbb{V}\mathrm{ar}[\hat{\theta}_a \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right] \\
&= \mathbb{E}\left[\frac{1}{\left(\sum_{i=1}^n \mathbb{1}(A_i=1)\right)^2} \sum_{i=1}^n \mathbb{V}\mathrm{ar}[h(Y_i) \mid \mathbf{A}]\big(\mathbb{1}(A_i=1)\big)^2 \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right] \\
&= \mathbb{E}\left[\frac{1}{\left(\sum_{i=1}^n \mathbb{1}(A_i=1)\right)^2} \sum_{i=1}^n \mathbb{V}\mathrm{ar}[h(Y_i) \mid A_i]\mathbb{1}(A_i=1) \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right] \\
&= \sum_{\substack{(a_1,\ldots,a_n)\in\{0,1\}^n \\ \sum_{i=1}^n a_i>0}} \Bigg\{\frac{1}{\left(\sum_{i=1}^n \mathbb{1}(a_i=1)\right)^2} \sum_{i=1}^n \mathbb{V}\mathrm{ar}[h(Y_i) \mid A_i=a_i]\mathbb{1}(a_i=1) \\
&\qquad\qquad \times \mathbb{P}\left(\mathbf{A}=(a_1,\ldots,a_n) \mid \sum_{i=1}^n a_i>0\right)\Bigg\} \\
&= \sum_{k=1}^n \binom{n}{k}\frac{1}{k^2} \times k\mathbb{V}\mathrm{ar}[h(Y) \mid A=1]\left(\frac{1}{M}\right)^k\left(1-\frac{1}{M}\right)^{n-k}\left[1-\left(1-\frac{1}{M}\right)^n\right]^{-1} \\
&= \mathbb{V}\mathrm{ar}[h(Y) \mid A=1]\sum_{k=1}^n \binom{n}{k}\frac{1}{k}\left(\frac{1}{M}\right)^k\left(1-\frac{1}{M}\right)^{n-k}\left[1-\left(1-\frac{1}{M}\right)^n\right]^{-1} \\
&= \mathbb{V}\mathrm{ar}[h(Y) \mid A=1]\mathbb{E}\left[\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i=1)} \mid \sum_{i=1}^n \mathbb{1}(A_i=1)>0\right]
\end{aligned}
$$

où la dernière égalité vient du fait que

$$\sum_{i=1}^n \mathbb{1}(A_i=1) \sim \mathrm{Bin}\left(n, \frac{1}{M}\right)$$

Calculons maintenant $\mathbb{E}\big[\hat{\theta}_r \mid \sum_{i=1}^n \mathbb{1}(A_i=0)>0\big]$ et $\mathbb{V}\mathrm{ar}\big[\hat{\theta}_r \mid \sum_{i=1}^n \mathbb{1}(A_i=0)>0\big]$. Posons

$$T = \frac{1}{\sum_{i=1}^n \mathbb{1}(A_i=0)} \sum_{i=1}^n h(Y_i)\mathbb{1}(A_i=0)$$

On a

$$
\begin{aligned}
\mathbb{E}\left[T \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] &= \mathbb{E}\left[\mathbb{E}[T \mid \mathbf{A}] \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] \\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^{n} \mathbb{1}(A_i = 0)} \sum_{i=1}^{n} \mathbb{E}[h(Y_i) \mid \mathbf{A}] \mathbb{1}(A_i = 0) \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] \\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^{n} \mathbb{1}(A_i = 0)} \sum_{i=1}^{n} \mathbb{E}[h(Y_i) \mid A_i] \mathbb{1}(A_i = 0) \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] \\
&= \sum_{\substack{(a_1,\ldots,a_n) \in \{0,1\}^n \\ \sum_{i=1}^{n} a_i < n}} \Bigg\{ \frac{1}{\sum_{i=1}^{n} \mathbb{1}(a_i = 0)} \sum_{i=1}^{n} \mathbb{E}[h(Y_i) \mid A_i = a_i] \mathbb{1}(a_i = 0) \\
&\qquad \times \mathbb{P}\left(A = (a_1, \ldots, a_n) \mid \sum_{i=1}^{n} a_i < n\right) \Bigg\} \\
&= \sum_{k=1}^{n} \binom{n}{k} \frac{1}{k} \times k \mathbb{E}[h(Y) \mid A = 0] \times \left(\frac{1}{M}\right)^{n-k} \left(1 - \frac{1}{M}\right)^{k} \left[1 - \left(\frac{1}{M}\right)^{n}\right]^{-1} \\
&= \mathbb{E}[h(Y) \mid A = 0] \left[1 - \left(\frac{1}{M}\right)^{n}\right]^{-1} \sum_{k=1}^{n} \binom{n}{k} \left(\frac{1}{M}\right)^{n-k} \left(1 - \frac{1}{M}\right)^{k} \\
&= \mathbb{E}[h(Y) \mid A = 0] \left[1 - \left(\frac{1}{M}\right)^{n}\right]^{-1} \left[1 - \left(\frac{1}{M}\right)^{n}\right] \\
&= \mathbb{E}[h(Y) \mid A = 0]
\end{aligned}
$$

D'où

$$
\begin{aligned}
\mathbb{E}\left[\hat{\theta}_r \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] &= M\mathbb{E}[h(Y)] - (M-1)\mathbb{E}\left[T \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] \\
&= M\mathbb{E}[h(Y)] - (M-1)\mathbb{E}[h(Y) \mid A = 0] \\
&= \theta
\end{aligned}
$$

$\therefore$ Sachant l'événement $\left\{ \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0 \right\}$, l'estimateur $\hat{\theta}_r$ est sans biais pour θ. Pour la variance, on procède comme ce qui a été fait plus tôt pour l'estimateur $\hat{\theta}_a$. D'abord, par un argument tout à fait similaire à celui qui a été utilisé précédemment pour $\hat{\theta}_a$, on note que

$$
\mathbb{V}\mathrm{ar}\left[\mathbb{E}[\hat{\theta}_r \mid \mathbf{A}] \mid \sum_{i=1}^{n} \mathbb{1}(A_i = 0) > 0\right] = 0
$$

Ainsi,

$$\begin{aligned}
\mathbb{V}\text{ar}\Bigg[\hat{\theta}_r \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] &= \mathbb{E}\Bigg[\mathbb{V}\text{ar}[\hat{\theta}_r \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] + \mathbb{V}\text{ar}\Bigg[\mathbb{E}[\hat{\theta}_r \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{V}\text{ar}[\hat{\theta}_r \mid \mathbf{A}] \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] \\
&= (M-1)^2 \mathbb{E}\Bigg[\frac{1}{\big(\sum_{i=1}^n \mathbb{1}(A_i = 0)\big)^2} \sum_{i=1}^n \mathbb{V}\text{ar}[h(Y_i) \mid \mathbf{A}]\big(\mathbb{1}(A_i = 0)\big)^2 \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] \\
&= (M-1)^2 \mathbb{E}\Bigg[\frac{1}{\big(\sum_{i=1}^n \mathbb{1}(A_i = 0)\big)^2} \sum_{i=1}^n \mathbb{V}\text{ar}[h(Y_i) \mid A_i]\mathbb{1}(A_i = 0) \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] \\
&= (M-1)^2 \sum_{\substack{(a_1,\ldots,a_n)\in\{0,1\}^n \\ \sum_{i=1}^n a_i < n}} \Bigg\{\frac{1}{\big(\sum_{i=1}^n \mathbb{1}(a_i = 0)\big)^2} \sum_{i=1}^n \mathbb{V}\text{ar}[h(Y_i) \mid A_i = a_i]\mathbb{1}(a_i = 0) \\
&\qquad\qquad \times \mathbb{P}\Bigg(\mathbf{A} = (a_1,\ldots,a_n) \mid \sum_{i=1}^n a_i < n\Bigg)\Bigg\} \\
&= (M-1)^2 \sum_{k=1}^n \binom{n}{k}\frac{1}{k^2} \times k\mathbb{V}\text{ar}[h(Y) \mid A = 0]\bigg(\frac{1}{M}\bigg)^{n-k}\bigg(1 - \frac{1}{M}\bigg)^k \Bigg[1 - \bigg(\frac{1}{M}\bigg)^n\Bigg]^{-1} \\
&= (M-1)^2 \mathbb{V}\text{ar}[h(Y) \mid A = 0] \sum_{k=1}^n \binom{n}{k}\frac{1}{k}\bigg(\frac{1}{M}\bigg)^{n-k}\bigg(1 - \frac{1}{M}\bigg)^k \Bigg[1 - \bigg(\frac{1}{M}\bigg)^n\Bigg]^{-1} \\
&= (M-1)^2 \mathbb{V}\text{ar}[h(Y) \mid A = 0]\mathbb{E}\Bigg[\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i = 0)} \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg]
\end{aligned}$$

c) Lorsque M et n sont très grands, on a que

$$\mathbb{V}\text{ar}\Bigg[\hat{\theta}_a \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\Bigg] \approx \frac{M}{n}\mathbb{V}\text{ar}[h(Y) \mid A = 1]$$

et

$$\mathbb{V}\text{ar}\Bigg[\hat{\theta}_r \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Bigg] \approx \frac{(M-1)^2}{n}\mathbb{V}\text{ar}[h(Y) \mid A = 0]$$

par les approximations données dans l'énoncé. Ainsi,

$$\frac{\mathbb{V}\text{ar}\Big[\hat{\theta}_a \mid \sum_{i=1}^n \mathbb{1}(A_i = 1) > 0\Big]}{\mathbb{V}\text{ar}\Big[\hat{\theta}_r \mid \sum_{i=1}^n \mathbb{1}(A_i = 0) > 0\Big]} \approx K \times \frac{M}{(M-1)^2} \overset{M\to\infty}{\longrightarrow} 0 \tag{4}$$

où $0 < K < \infty$ est le rapport des variances. On constate donc que $\hat{\theta}_a$ est plus efficace que $\hat{\theta}_r$.
Voici quelques observations:

- L'espérance du nombre d'observations acceptées est

$$\mathbb{E}\Bigg[\sum_{i=1}^n \mathbb{1}(A_i = 1)\Bigg] = \frac{n}{M}$$

De plus,

$$\mathbb{V}\text{ar}[h(Y) \mid A = 1] = \mathbb{V}\text{ar}[h(X)]$$

On peut donc interpréter l'approximation

$$\frac{M}{n}\mathbb{V}\text{ar}[h(Y) \mid A = 1]$$

comme étant la variance de la loi dont on cherche à estimer l'espérance (à savoir la loi de la variable aléatoire $h(X)$) divisée par la taille échantillonnale. Cela correspond à la variance bien connue de l'estimateur canonique de la moyenne d'une loi, à savoir la moyenne empirique. Ceci n'est pas surprenant dans la mesure où θ_a estime θ *directement*. En effet, cet estimateur est simplement la moyenne empirique des observations acceptées auxquelles la fonction h a été appliquée. On s'attend donc à ce que cet estimateur soit précis (ait une faible variance) si la taille échantillonnale $\left(\frac{n}{M}\right)$ est grande.

- En revanche, $\hat{\theta}_r$ estime θ *indirectement* en corrigeant, d'une certaine manière, $\mathbb{E}[h(Y)]$. En effet, on peut réécrire

$$\hat{\theta}_r = \mathbb{E}[h(Y)] + (M-1)\left(\mathbb{E}[h(Y)] - \frac{1}{\sum_{i=1}^n \mathbb{1}(A_i = 0)}\sum_{i=1}^n h(Y_i)\mathbb{1}(A_i = 0)\right) \tag{5}$$

Si M est très grand, alors très peu d'observations sont acceptées (par rapport au nombre total de simulations effectuées). Ainsi, la loi de la variable aléatoire

$$\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i = 0)}\sum_{i=1}^n h(Y_i)\mathbb{1}(A_i = 0)$$

est très proche de celle de

$$\frac{1}{n}\sum_{i=1}^n h(Y_i)$$

Si n est grand, alors $\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i=0)}\sum_{i=1}^n h(Y_i)\mathbb{1}(A_i = 0)$ devrait prendre des valeurs proches de son espérance (par la loi forte des grands nombres) qui est elle-même proche, sans être égale, à $\mathbb{E}[h(Y)]$. On s'attend donc à ce que la différence entre parenthèses dans (5) soit très petite. Par contre, cette différence est multipliée par le facteur $(M-1)$, ce qui explique (intuitivement) pourquoi sa variance est plus élevée que celle de $\hat{\theta}_a$. En d'autres termes, $\hat{\theta}_r$ est un estimateur moins efficace que $\hat{\theta}_a$ pour estimer $\theta = \mathbb{E}[h(X)]$, car il corrige une certaine quantité connue ($\mathbb{E}[h(Y)]$) en moyennant des observations qui sont tirées d'une loi – à savoir celle de $\frac{1}{\sum_{i=1}^n \mathbb{1}(A_i=0)}\sum_{i=1}^n h(Y_i)\mathbb{1}(A_i = 0)$– qui n'est pas celle dont on veut estimer l'espérance – à savoir celle de $h(X)$. L'estimateur $\hat{\theta}_a$, quant à lui, estime θ en moyennant des observations dont l'espérance est θ.

d) Les résultats obtenus sont cohérents avec ce qui a été expliqué en c). Les moyennes estimées des deux estimateurs sont très proches du $\theta = \frac{2}{7}$, ce qui suggère qu'ils sont tout deux non biaisés. De plus, la variance de $\hat{\theta}_a$ est beaucoup plus petite que celle de $\hat{\theta}_r$. En fait, le rapport des variances est environ $9.2957 \times 10^{-4} \approx 0$, ce qui est cohérent avec (4).