Aix*Marseille université

Master 2 informatique

Parcours IAAA, UE MIA

Feuille d'exercices numéro 4

Année 2025-26

Calculs de moments et statistique.

Exercice 1 Preuve de propriétés vues en cours. ($\star$)

1. Montrez que $\mathrm{Var}(X) = E[X^2] - E[X]^2$
2. Prouvez la décomposition biais-variance pour la fonction de perte quadratique
3. Donnez une expression pour la variance de la somme de variables aléatoires (pas forcément indépendantes) en fonction des variances et covariances des variables impliquées dans cette somme.
4. Prouvez la loi de l'espérance totale.
5. Déduisez la loi de la variance totale de la loi de la covariance totale.
6. Prouvez la loi de la variance totale.

Exercice 2 Espérance et variance d'estimateurs classiques. ($\star$)

Soit n un entier naturel et $X_1, X_2, \ldots, X_n$ des variables aléatoires réelles mutuellement indépendantes. Par souci de simplicité, on suppose que tous les moments des ces variables aléatoires existent. On note

$$\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

et

$$\hat{V} := \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$$

1. Montrez que

$$\hat{V} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \hat{\mu}^2$$

Solution : On développe les carrés et on regroupe les termes en trois sommes, puis on factorise et on reconnaît la définition de la moyenne empirique :

$$\begin{aligned}\hat{V} &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} 2\hat{\mu}X_i + \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\hat{\mu}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) + \hat{\mu}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\hat{\mu}\hat{\mu} + \hat{\mu}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \hat{\mu}^2.\end{aligned}$$

2. Exprimez l'espérance de $\hat{\mu}$ comme une fonction de l'espérance et des moments centrés des variables $X_1, X_2, \ldots, X_n$ (prouvez votre résultat)

Solution : Notons $\mu_i := \mathbf{E}[X_i]$ pour $i = 1 \ldots n$. Par linéarité de l'ésperance :

$$\mathbf{E}[\hat{\mu}] := \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu_i.$$

On suppose à présent que $X_1, X_2, \ldots, X_n$ suivent toute la même distribution (elles sont donc i.i.d. puisqu'on a déjà supposé qu'elles étaient indépendantes).

3. Répondez à nouveau à la question précédente dans ce cadre plus simple.

Solution : Notons $\mu := \mathbf{E}[X_i]$ pour $i = 1 \ldots n$.

On a à présent :

$$\mathbf{E}[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

4. Exprimez la variance de $\hat{\mu}$ comme une fonction de l'espérance et des moments centrés des variables $X_1, X_2, \ldots, X_n$ (prouvez votre résultat)

Solution : Notons $\sigma^2 := \mathbf{Var}[X_i]$ pour $i = 1 \dots n$. On a :

$$\begin{aligned}\mathbf{Var}[\hat{\mu}] &= \mathbf{E}[\hat{\mu}^2] - \mathbf{E}[\hat{\mu}]^2 \\ &= \mathbf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] - \mu^2 \\ &= \mathbf{E}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] - \mu^2.\end{aligned}$$

Par linéarité de l'espérance, on obtient :

$$\mathbf{Var}[\hat{\mu}] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{E}[X_i X_j] - \mu^2.$$

Pour $i \neq j$, les variables X_i et X_j sont indépendantes et on obtient $\mathbf{E}[X_i X_j] = \mathbf{E}[X_i]\mathbf{E}[X_j] = \mu^2$. Pour $i = j$, on a : $\mathbf{E}[X_i X_j] = \mathbf{E}[X_i]^2 = \sigma^2 + \mu^2$. Au final :

$$\begin{aligned}\mathbf{Var}[\hat{\mu}] &= \frac{1}{n^2}\left(n(\sigma^2 + \mu^2) + n(n-1)\mu^2\right) - \mu^2 \\ &= \frac{\sigma^2}{n}.\end{aligned}$$

5. Exprimez l'espérance de $\hat{V}$ comme une fonction de l'espérance et des moments centrés des variables $X_1, X_2, \dots, X_n$ (prouvez votre résultat)

Solution : Par linéarité de l'espérance :

$$\begin{aligned}\mathbf{E}[\hat{V}] &= \mathbf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \hat{\mu}^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}[X_i^2] - \mathbf{E}[\hat{\mu}^2].\end{aligned}$$

Or on a vu dans la réponse à la question précédente que $\mathbf{E}[X_i^2] = \mu^2 + \sigma^2$ et que $\mathbf{E}[\hat{\mu}^2] = \mu^2 + \sigma^2/n$. Donc :

$$\mathbf{E}[\hat{V}] = \sigma^2(1 - 1/n) = \frac{n-1}{n}\sigma^2.$$

6. Exprimez la variance de $\hat{V}$ comme une fonction de l'espérance et des moments centrés des variables $X_1, X_2, \dots, X_n$ (prouvez votre résultat)

Exercice 3 Loi de l'espérance totale. ($\star$)

Un téléphone kadok tient en moyenne $12h$ avec la batterie A, mais seulement $8h$ avec la batterie B. La batterie A se trouve dans 80% des téléphones kadok, le reste étant muni de la batterie B. Si vous achetez un téléphone kadok, combien d'heure vous attendez-vous à ce qu'il tienne ?

Exercice 4 Calculs de biais, variance, risque. ($\star$)

1. Calculer le biais et la variance de l'estimateur de la moyenne empirique pour un échantillon i.i.d.

Solution : Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires i.i.d. de moyenne μ et de variance σ^2. L'estimateur de la moyenne empirique est donné par :

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Biais :

$$\text{Biais}(\hat{\mu}) = \mathbb{E}[\hat{\mu}] - \mu = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] - \mu = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] - \mu = \mu - \mu = 0.$$

Donc, l'estimateur est non biaisé.

Variance :

$$\text{Var}(\hat{\mu}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) + 0 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

2. Calculer le biais de l'estimateur de la variance suivant pour un échantillon i.i.d. : $\hat{V} := \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$

Solution : L'estimateur $\hat{V}$ est l'estimateur de la variance avec le dénominateur n. Nous cherchons $\text{Biais}(\hat{V}) = \mathbb{E}[\hat{V}] - \sigma^2$.

Calcul de $\mathbb{E}[\hat{V}]$:

$$\mathbb{E}[\hat{V}] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(X_i - \hat{\mu})^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^2] - 2\mathbb{E}[\hat{\mu}X_i] + \mathbb{E}[\hat{\mu}^2]$$

Or

(a) $\sigma^2 = \text{Var}[X_i] = \mathbb{E}[X_i^2] - \mu^2$, donc $\mathbb{E}[X_i^2] = \mu^2 + \sigma^2$.

(b) $\mathbb{E}[\hat{\mu}X_i] = \frac{1}{n}\sum_j \mathbb{E}[X_i X_j] = \frac{n-1}{n}\mu^2 + \frac{1}{n}\mathbb{E}[X_i^2] = \frac{n-1}{n}\mu^2 + \frac{\sigma^2}{n} + \frac{\mu^2}{n} = \mu^2 + \frac{\sigma^2}{n}$.

(c) $\mathbb{E}[\hat{\mu}^2] = \frac{1}{n^2}\sum_i\sum_j \mathbb{E}[X_i X_j] = \frac{1}{n^2}\sum_i \mathbb{E}[X_i^2] + \frac{1}{n^2}\sum_{i\neq j}\mu^2$
$= \frac{n}{n^2}(\mu^2 + \sigma^2) + \frac{n(n-1)}{n^2}\mu^2 = \mu^2 + \frac{\sigma^2}{n}$.

Donc

$$\mathbb{E}[\hat{V}] = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 - \frac{\sigma^2}{n}\right) = \sigma^2\left(1 - \frac{1}{n}\right).$$

Biais :

$$\text{Biais}(\hat{V}) = \mathbb{E}[\hat{V}] - \sigma^2 = \sigma^2\left(1 - \frac{1}{n}\right) - \sigma^2 = -\frac{\sigma^2}{n}.$$

Donc, l'estimateur est biaisé vers le bas de $-\dfrac{\sigma^2}{n}$.

3. Proposez un estimateur non-biaisé de la variance pour un échantillon i.i.d.

Solution : Un estimateur non biaisé de la variance est obtenu en utilisant le dénominateur $n-1$:

$$\tilde{V} := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2.$$

Vérification du caractère non biaisé :

$$\mathbb{E}[\tilde{V}] = \mathbb{E}\left[\frac{n-1}{n}\hat{V}\right] = \frac{n-1}{n}\mathbb{E}[\hat{V}] = \sigma^2.$$

Ainsi, $\tilde{V}$ est un estimateur non biaisé de la variance.

4. Comparez le risque quadratique moyen des deux estimateurs de la variance.

5. Pouvez-vous donner un estimateur avec un risque plus faible que les deux considérés jusqu'ici ?

Exercice 5 Analyse des propriétés basiques d'un estimateur. ($\star\star$)

Supposons qu'on observe un échantillon $x_1, x_2, \ldots, x_n \in (\mathbf{R}^d)^n$, i.i.d. de distribution P. Soit $d : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ une fonction mesurant la 'dissimilarité' entre deux points de $\mathbf{R}^d$. On suppose que d est symmétrique, c'est à dire que $d(x_1, x_2) = d(x_2, x_1)$ pour tout choix de x_1, x_2. Nous cherchons à estimer la dissimilarité moyenne entre deux points tirés aléatoirement et indépendamment suivant la distribution P :

$$\delta(P, d) := \mathbf{E}_{a,b\sim P\otimes P}[d(a, b)]$$

sur la base de $x_1, x_2, ..., x_n$ (la notation $a, b \sim P \otimes P$ signifie que a et b sont deux échantillons tirés indépendamment de la loi P). On suppose que $\mathbf{E}_{a,b\sim P\otimes P}[d(a, b)^2] < +\infty$.

Considérons l'estimateur :

$$\hat{\delta}(x_1, x_2, \ldots, x_n) := \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} d(x_i, x_j).$$

1. Quel est le biais de $\hat{\delta}$?

Solution : Par linéarité de l'espérance :

$$\begin{aligned} b(\hat{\delta}) &:= \mathbf{E}[\hat{\delta}] - \delta \\ &= \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbf{E}[d(x_i, x_j)] - \delta \end{aligned}$$

Comme on a toujours $i \neq j$ et que $x_1, \ldots, x_n$ sont i.i.d de loi P, on a toujours $\mathbf{E}[d(x_i, x_j)] = \delta$. Donc :

$$\begin{aligned} b(\hat{\delta}) &= \left(\frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \delta \right) - \delta \\ &= \delta - \delta \\ &= 0. \end{aligned}$$

$\hat{\delta}$ est donc un estimateur non biaisé de δ.

2. Quelle est la variance de $\hat{\delta}$? On l'exprimera en fonction de $\sigma_1^2 = \mathrm{Var}_{x_1\sim P}\mathbf{E}_{x_2\sim P}[d(x_1, x_2)]$ et $\sigma_2^2 = \mathrm{Var}_{(x_1,x_2)\sim P\otimes P}[d(x_1, x_2)]$.

Solution : On peut, par exemple, appliquer la formules donnant la variance d'une quantité multipliée par une constante et la formule donnant la variance d'une somme en fonction des covariances des termes:

$$\mathbf{Var}(\hat{\delta}) := \frac{1}{\binom{n}{2}^2} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \sum_{k=1}^{n} \sum_{l=k+1}^{n} \mathbf{Cov}(d(x_i, x_j), d(x_k, x_l))$$

On distingue trois cas :

- Si $\{i, j\} \cap \{k, l\} = \emptyset$, alors $\mathbf{Cov}(d(x_i, x_j), d(x_k, x_l)) = 0$.
- Sinon, si $|\{i, j\} \cap \{k, l\}| = 1$, alors par symétrie de d,

$$\begin{aligned}
\mathbf{Cov}(d(x_i, x_j), d(x_k, x_l)) &= \mathbf{Cov}_{a,b,c\sim P\otimes P\otimes P}(d(a, b), d(a, c)) \\
&= \mathbf{E}_{a,b,c\sim P\otimes P\otimes P}(d(a, b), d(a, c)) - [\mathbf{E}_{a,b\sim P\otimes P}(d(a, b))]^2 \\
&= \mathbf{E}_{a\sim P}[\mathbf{E}_{b,c\sim P\otimes P}[(d(a, b), d(a, c))]] - [\mathbf{E}_{a,b\sim P\otimes P}(d(a, b))]^2 \\
&= \mathbf{E}_{a\sim P}[\mathbf{E}_{b\sim P}[(d(a, b)]\mathbf{E}_{c\sim P}[d(a, c))]] - [\mathbf{E}_{a,b\sim P\otimes P}(d(a, b))]^2 \\
&= \mathbf{E}_{a\sim P}[\mathbf{E}_{b\sim P}[(d(a, b)]^2] - [\mathbf{E}_{a,b\sim P\otimes P}(d(a, b))]^2 \\
&= \sigma_1^2.
\end{aligned}$$

- Sinon, $\{i, j\} = \{k, l\}$ et $\mathbf{Cov}(d(x_i, x_j), d(x_k, x_l)) = \mathbf{Var}(d(x_i, x_j)) = \sigma_2^2$.

De plus, $\{i, j\} = \{k, l\}$ pour exactement $\binom{n}{2}$ quadruplets parmi les $\binom{n}{2}^2$ quadruplets considérés et $|\{i, j\} \cap \{k, l\}| = 1$ pour exactement :

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} n - i - 1 + n - j + i - 1 + j - 2 = 2(n - 2)\binom{n}{2}$$

quadruplets parmi les $\binom{n}{2}^2$ quadruplets considérés (pour chaque choix de i et j on a $n - i - 1$ cas avec $k = i$ et $l \neq j$, $n - j$ cas avec $k = j$ et $l \neq i$, $i - 1$ cas avec $l = i$ et $k \neq j$ et $j - 2$ cas avec $l = j$ et $k \neq i$).
Au final, on obtient :

$$\mathbf{Var}(\hat{\delta}) := \binom{n}{2}^{-1} [2(n - 2)\sigma_1^2 + \sigma_2^2].$$

3. Prouver l'inégalité de Markov: soit X est un variable aléatoire réelle positive, de moyenne finie μ et soit t un réel strictement positif, alors :

$$p(X \geq t) \leq \frac{\mu}{t}.$$

Solution : Notons P la loi de X. Par définition :

$$\mu = \mathbf{E}[X] = \int X dP$$

Soit $t > 0$. Comme X est positive, on a toujours : $X \geq \mathbf{1}_{X \geq t} t$, où $\mathbf{1}_{X \geq t}$ est la fonction qui vaut 0 si $X < t$ et 1 sinon. On en déduit par monotonie de l'espérance que :

$$\mu \geq \int \mathbf{1}_{X \geq t} t dP = t \int \mathbf{1}_{X \geq t} dP = t p(X \geq t)$$

Donc :

$$p(X \geq t) \leq \frac{\mu}{t}.$$

4. Utiliser l'inégalité de Markov pour prouver l'inégalité de Chebyshev: soit X est un variable aléatoire réelle de moyenne finie μ et de variance finie σ^2 et soit t un réel strictement positif, alors :

$$p(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Solution : On considére $Y = (X - \mu)^2$. Y est positive et d'espérance égale à la variance de X (par définition), qui est finie et égale à σ^2 par hypothèse. Soit $t > 0$. $t^2 > 0$, on peut donc appliquer l'inégalité de Markov à Y en t^2. On obtient :

$$p((X - \mu)^2 \geq t^2) \leq \frac{\sigma^2}{t^2}.$$

Or $(X - \mu)^2 \geq t^2$ est équivalent à $|X - \mu| \geq t$, donc les évènements $\{(X - \mu)^2 \geq t^2\}$ et $\{|X - \mu| \geq t\}$ sont identiques et $p((X - \mu)^2 \geq t^2) = p(|X - \mu| \geq t)$. On obtient donc finalement :

$$p(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

5. Utiliser l'inégalité de Chebyshev et les résultats des deux premières questions pour montrer que $\hat{\delta}$ est un estimateur *faiblement consistant* de δ, c'est à dire qu'on a $\hat{\delta} \rightarrow_p \delta$, c'est à dire

que pour tout $\epsilon > 0$:

$$\lim_{n \to +\infty} p(|\hat{\delta}(x_1, \dots x_n) - \delta| \geq \epsilon) = 0.$$

Solution : $\hat{\delta}$ est un estimateur faiblement consistant de δ si et seulement si on a $\hat{\delta} \to_p \delta$, c'est à dire que pour tout $\epsilon > 0$:

$$\lim_{n \to +\infty} p(|\hat{\delta}(x_1, \dots x_n) - \delta| \geq \epsilon) = 0.$$

Soit $\epsilon > 0$ et n un entier supérieur ou égal à un. $\hat{\delta}(x_1, \dots, x_n)$ vérifie les hypothèses d'application pour l'inégalité de Chebyshev, qu'on applique en $t = \epsilon$ pour obtenir (en utilisant les résultats des deux premières questions pour exprimer l'espérance et la variance de $\hat{\delta}$) :

$$p((|\hat{\delta}(x_1, \dots x_n) - \delta| \geq \epsilon) \leq \frac{\binom{n}{2}^{-1}[2(n-2)\sigma_1^2 + \sigma_2^2]}{\epsilon^2}.$$

Le membre de droite de cette inégalité tend vers 0 quand n tend vers $+\infty$ (il est en $O(1/n)$) et le membre de gauche est positif (puisqu'il s'agit d'une probabilité). On peut donc appliquer le théorème des gendarmes pour conclure que :

$$\lim_{n \to +\infty} p(|\hat{\delta}(x_1, \dots x_n) - \delta| \geq \epsilon) = 0.$$

Exercice 6 On considère des variables aléatoires réelles $X_1, X_2, \dots, X_n$ identiquement distribuées de loi P, *mais pas forcément indépendantes.* On suppose que l'espérance et la variance de P sont finies et on note $\mu = E[P] \in \mathbf{R}$, $\sigma^2 = \mathrm{Var}[P] \in \mathbf{R}$. On suppose, de plus, qu'il existe $\alpha > 0$ tel que pour tout $1 \leq i \leq n$ et $1 \leq j \leq n$, $\mathrm{Cov}(X_i, X_j) = \frac{1}{2^{\alpha|i-j|}}$. (On pourrait rencontrer cette situation, par exemple, dans le cadre d'un échantillonnage uniforme d'un signal aléatoire de loi stationnaire.)

On s'intéresse au propriété statistiques de l'estimateur de la moyenne empirique :

$$\hat{\mu}(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

1. Calculez le biais de $\hat{\mu}$.
2. Calculez la variance de $\hat{\mu}$. Vous pouvez utiliser directement les formules suivantes: pour tout $q < 1$, $\sum_{i=0}^{n} u_0 q^i = u_0 \frac{1-q^{n+1}}{1-q}$ et $\sum_{i=0}^{n} u_0 i q^i = u_0 \frac{q(nq^{n+1} - (n+1)q^n + 1)}{(1-q)^2}$
3. Que se passe-t-il quand $\alpha \to +\infty$ à la question précédente? (Calculez cette limite et proposez une interprétation).

4. Même question pour $\alpha \to 0$.

Exercice 7 On considère des variables aléatoires réelles $X_1, X_2, \dots X_n$ identiquement distribuées de loi P, *mais pas forcément indépendantes*. On note $X := [X_1|X_2|\dots|X_n]$ la matrice dont les colonnes contiennent ces variables. On suppose que tous les moments nécessaires sont finis, on note $\mu = E[P] \in \mathbf{R}$, $\sigma^2 = \mathrm{Var}[P] \in \mathbf{R}$ et on note $\Sigma = (c_{i,j}) \in \mathbf{R}^{n\times n}$ la matrice de covariance (théorique) des n variables aléatoires $X_1, X_2, ..., X_n$.

1. Que pouvez-vous dire sur $c_{i,i}$?

   Solution : $c_{i,i} = \mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i) = \sigma_2$

2. Calculez l'espérance μ_Σ et la variance V_Σ de $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i$ en fonction des coefficients $c_{i,j}$ et de μ.

   Solution : Indications de correction.

   En utilisant la linéarité de l'espérance (qui ne nécessite pas d'indépendance entre les variables), on trouve $\mu_\Sigma = \mu$, comme dans l'exercice 4.

   En utilisant la bilinéarité de la covariane et $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$, on obtient : $V_\Sigma = \frac{1}{n^2}\sum_i \sum_j c_{i,j} = \frac{\sigma^2}{n} + \frac{2}{n^2}\sum_i \sum_{j>i} c_{i,j}$

3. Pour un σ fixé, quel choix de Σ maximise la variance de $\hat{\mu}$? Quel choix minimise sa variance ?

   Solution : Indications de correction.

   On peut avoir l'intuition que la variance sera maximale quand toutes les variables sont "maximalement" dépendantes, c'est à dire $c_{i,j} = \sigma^2$. Pour le démontrer et pour montrer qu'il existe un choix de $c_{i,j}$ qui donne une variance de zéro, une bonne méthode est de commencer par considérer une décomposition en éléments propre de Σ et d'exprimer V_Σ à l'aide de ces éléments propres.

Exercice 8 Soit (X, Y) un vecteur aléatoire de $\mathbf{R}^2$ de distribution jointe P. On suppose que les moments suivants de P existent: $E[(X, Y)] = (\mu_X, \mu_Y)$, $E[\mathrm{Var}[X|Y]] = \sigma_1^2$, $\mathrm{Var}[E[X|Y]] = \sigma_2^2$.

On note P_Y la loi marginale de Y et $P_{X|Y=y}$ la loi conditionnelle de X sachant $Y = y$.

Considérons le processus d'échantillonnage stratifié suivant:

- Y_1, Y_2 forment un échantillon i.i.d. de loi P_Y
- X_{11}, X_{12} forment un échantillon i.i.d. de loi $P_{X|Y=Y_1}$
- X_{21}, X_{22}, X_{23} un échantillon i.i.d. de loi $P_{X|Y=Y_2}$.

On s'intéresse au propriétés statistiques des estimateurs :

$$\hat{\mu}_1(X_{11}, X_{12}, X_{21}, X_{22}, X_{23}) = \frac{1}{5}(X_{11} + X_{12} + X_{21} + X_{22} + X_{23})$$

et

$$\hat{\mu}_2(X_{11}, X_{12}, X_{21}, X_{22}, X_{23}) = \frac{1}{2}\left(\frac{1}{2}(X_{11} + X_{12}) + \frac{1}{3}(X_{21} + X_{22} + X_{23})\right)$$

1. Calculez l'espérance de $\hat{\mu}_1$ et $\hat{\mu}_2$ et commentez.

Solution : Pour tout i, j,

$$\mathbf{E}[X_{ij}] = \mu_X$$

donc par linéarité de l'espérance,

$$\mathbf{E}[\hat{\mu}_1] = \mathbf{E}[\hat{\mu}_2] = \mu_X$$

Les deux estimateurs sont donc deux estimateurs différents de μ_X, on peut se demander si l'un est plus efficace que l'autre, ce qui nous amène à la question suivante.

2. Calculez la variance de $\hat{\mu}_1$ et $\hat{\mu}_2$ et commentez

Solution : Faisons le calcul en détail par exemple pour $\hat{\mu}_2$. On utilise la bilinéarité de la covariance :

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_2) = \mathrm{Covar}(\hat{\mu}_2, \hat{\mu}_2) = &\frac{1}{16}\mathrm{Covar}(X_{11}, X_{11}) + \frac{1}{16}\mathrm{Covar}(X_{11}, X_{12}) + \\ &\frac{1}{16}\mathrm{Covar}(X_{12}, X_{11}) + \frac{1}{16}\mathrm{Covar}(X_{12}, X_{12}) + \\ &\frac{1}{36}\mathrm{Covar}(X_{21}, X_{21}) + \frac{1}{36}\mathrm{Covar}(X_{21}, X_{22}) + \frac{1}{36}\mathrm{Covar}(X_{21}, X_{23}) + \\ &\frac{1}{36}\mathrm{Covar}(X_{22}, X_{21}) + \frac{1}{36}\mathrm{Covar}(X_{22}, X_{22}) + \frac{1}{36}\mathrm{Covar}(X_{22}, X_{23}) + \\ &\frac{1}{36}\mathrm{Covar}(X_{23}, X_{21}) + \frac{1}{36}\mathrm{Covar}(X_{23}, X_{22}) + \frac{1}{36}\mathrm{Covar}(X_{23}, X_{23}) + \\ &\frac{1}{24}\mathrm{Covar}(X_{11}, X_{21}) + \frac{1}{24}\mathrm{Covar}(X_{11}, X_{22}) + \frac{1}{24}\mathrm{Covar}(X_{11}, X_{23}) + \\ &\frac{1}{24}\mathrm{Covar}(X_{12}, X_{21}) + \frac{1}{24}\mathrm{Covar}(X_{12}, X_{22}) + \frac{1}{24}\mathrm{Covar}(X_{12}, X_{23})\end{aligned}$$

Dans la somme ci-dessus, on peut exprimer les différents termes en fonction de σ_1 et σ_2 en utilisant la loi de la (co)variance totale. Spécifiquement on a:

- Variance d'une variable :

$$\mathrm{Covar}(X_{ij}, X_{ij}) = \mathrm{Var}(X_{ij}) = \mathrm{Var}(\mathbf{E}[X_{ij} \mid Y_i]) + \mathbf{E}[\mathrm{Var}(X_{ij} \mid Y_i)] = \sigma_2^2 + \sigma_1^2$$

- Covariance intra-groupe ($j \neq k$):

$$\begin{aligned}\mathrm{Covar}(X_{ij}, X_{ik}) = &\quad \mathrm{Covar}(\mathbf{E}[X_{ij} \mid Y_i], \mathbf{E}[X_{ik} \mid Y_i]) + \quad \mathbf{E}[\mathrm{Covar}(X_{ij}, X_{ik} \mid Y_i)] \\ = &\quad \mathrm{Var}(\mathbf{E}[X \mid Y]) + \quad 0 \\ = &\quad \sigma_2^2\end{aligned}$$

- Covariance entre groupes ($i \neq k$):

$$\mathrm{Covar}(X_{ij}, X_{kl}) = 0$$

Au final on obtient :

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_2) = &\quad \left(2 \times \frac{1}{16} + 3 \times \frac{1}{36}\right)(\sigma_2^2 + \sigma_1^2) + \left(2 \times \frac{1}{16} + 6 \times \frac{1}{36}\right)\sigma_2^2 + 6 \times \frac{1}{24} \times 0 \\ = &\quad \frac{1}{2}\sigma_2^2 + \frac{5}{24}\sigma_1^2.\end{aligned}$$

Solution : (Solution continuée)

En suivant un raisonnement similaire, on obtient pour $\hat{\mu}_1$:

$$\mathrm{Var}(\hat{\mu}_1) = \frac{1}{25}(5(\sigma_2^2 + \sigma_1^2) + 8\sigma_2^2) = \frac{13}{25}\sigma_2^2 + \frac{1}{5}\sigma_1^2.$$

On voit que σ_2^2 a un poids plus fort dans la variance de $\hat{\mu}_1$ et σ_1^2 a un poids plus fort dans la variance de $\hat{\mu}_2$. En d'autres termes, aucun des deux estimateurs n'est strictement meilleur que l'autre, tout dépend des caractéristique de la loi P. Si la variance intra-groupe σ_1^2 domine, alors $\hat{\mu}_1$ aura une variance plus faible et sera donc préférable. Si la variance entre groupes σ_2^2 domine, alors $\hat{\mu}_2$ aura une variance plus faible et sera préférable.

Exercice 9 On entraîne des classifieur binaire d'images (par exemple pour la classification d'images de chiens vs chats) et on s'intéresse à la performance de la procédure d'entraînement plutôt qu'à la performance d'un classifieur entraîné particulier (par exemple pour comparer cette procédure à une autre). On veut également que nos résultats ne dépendent pas du choix particulier des images d'entraînement et de test. Comme on ne peut pas entraîner un grand nombre de modèles à cause du coût en temps de calcul de l'entraînement, on se limite à entraîner n modèles différents, chacun entraîné sur un sous-ensemble distinct d'images correspondant à un n-ième de l'ensemble d'entraînement disponible (par exemple un dixième si on entraîne dix modèles) et avec une initialisation aléatoire différente. On teste ensuite chacun de ces n modèles sur chacune des m images disponibles dans l'ensemble de test. On obtient ainsi une matrice $(e_{i,j})_{1\leq i\leq n,\ 1\leq j\leq m}$ d'erreurs de classification (chaque entrée $e_{i,j}$ de la matrice vaut soit 0, soit 1, selon que l'image j a été correctement classifiée par le classifieur i ou non).

1. Proposez un estimateur de l'espérance de l'erreur de classification prenant en compte tous les facteurs de variabilité considérés (choix des données d'entraînement et de test et stochasticité de la procédure d'entraînement).

2. Proposez un estimateur de la variance de l'erreur de classification prenant en compte tous les facteurs de variabilité considérés (choix des données d'entraînement et de test et stochasticité de la procédure d'entraînement).