Aix*Marseille université	Master 2 informatique Parcours IAAA, UE MIA
	Feuille d'exercices numéro 5
Année 2023-24	Statistiques.

Exercice 1 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.
2. Calculer le biais et la variance de l'estimateur de la variance suivant pour un échantillon i.i.d. : $\hat{V} := \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$
3. Trouvez un estimateur non-biaisé de la variance pour un échantillon i.i.d.
4. Prouvez la décomposition biais-variance pour le risque quadratique moyen
5. Comparez le risque quadratique moyen des deux estimateurs de la variance.
6. Pouvez-vous donner un estimateur avec un risque plus faible que les deux considérés jusqu'ici ?
7. On s'intéresse à présent au problème d'estimer la variance de l'estimateur de la moyenne empirique. Que pouvons-nous déduire des questions précédentes à ce sujet ?

Exercice 2 Estimation par maximum de vraisemblance des paramètres d'une loi Gaussienne multivariée. ($\star$)

Supposons qu'on observe un échantillon i.i.d. $x_1, x_2, \ldots, x_n \in (\mathbf{R}^d)^n$ de loi gaussienne multivariée $\mathcal{N}(\mu^*, \Sigma^*)$, pour un vecteur $\mu \in \mathbf{R}^d$ et une matrice $\Sigma^* \in \mathbf{S}_d$, où $\mathbf{S}_d$ est l'ensemble des matrices à coefficients réels symétriques définies positives de taille d pas d. On cherche à estimer μ^* et Σ^* à partir de l'observation de $x_1, x_2, \ldots, x_n$.

On définit la *vraisemblance* d'un couple de paramètres $(\mu, \Sigma) \in \mathbf{R}^d \times \mathbf{S}_d$ comme :

$$\ell(\mu, \Sigma) := p(x_1, x_2, \ldots, x_n; \mu, \Sigma),$$

où p corresponds à la densité de probabilité de $x_1, x_2, \ldots, x_n$.

On considère l'estimateur du *maximum de vraisemblance* pour μ^*, Σ^* défini par

$$\hat{\mu}, \hat{\Sigma} \in \arg \max_{(\mu,\Sigma)\in\mathbf{R}^d\times\mathbf{S}_d} \ell(\mu, \Sigma).$$

1. Montrer que

$$\hat{\mu}, \hat{\Sigma} \in \arg \max_{(\mu,\Sigma)\in \mathbf{R}^d \times \mathbf{S}_d} \log(\ell(\mu, \Sigma)).$$

Solution : Soit $\hat{\mu}, \hat{\Sigma} \in \arg\max_{(\mu,\Sigma)\in \mathbf{R}^d \times \mathbf{S}_d} \ell(\mu, \Sigma)$. Soit $(\mu, \Sigma) \in \mathbf{R}^d \times \mathbf{S}_d$. On a $\ell(\mu, \Sigma) \leq \ell(\hat{\mu}, \hat{\Sigma})$. Or log est une fonction croissante, donc $\log(\ell(\mu, \Sigma)) \leq \log(\ell(\hat{\mu}, \hat{\Sigma}))$. On en déduit que $\sup_{(\mu,\Sigma)\in \mathbf{R}^d \times \mathbf{S}_d} \log(\ell(\mu, \Sigma)) \leq \log(\ell(\hat{\mu}, \hat{\Sigma}))$ et comme $(\hat{\mu}, \hat{\Sigma}) \in \mathbf{R}^d \times \mathbf{S}_d$:

$$\hat{\mu}, \hat{\Sigma} \in \arg \max_{(\mu,\Sigma)\in \mathbf{R}^d \times \mathbf{S}_d} \log(\ell(\mu, \Sigma)).$$

2. Donner une expression (la plus simple que vous pouvez) pour $\log(\ell(\mu, \Sigma)))$ en utilisant la formule donnant la densité d'une loi gaussienne multivariée non dégénérée.

Solution : L'échantillon étant considéré i.i.d., on a :

$$\begin{aligned}
\log(\ell(\mu, \Sigma))) &= \log\left(\prod_{i=1}^{n} p(x_i; \mu, \Sigma)\right) \\
&= \sum_{i=1}^{n} \log\left(p(x_i; \mu, \Sigma)\right) \\
&= \sum_{i=1}^{n} \log\left(\frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)\right)\right) \\
&= -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \sum_{i=1}^{n} \frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu).
\end{aligned}$$

3. On suppose que la matrice de covariance empirique $\frac{1}{n}(x_i - \mu)(x_i - \mu)^T$ est inversible et on admet que $\log(\ell(\mu, \Sigma)))$ est de classe C^1 et admet un unique maximum sur $\mathbf{R}^d \times \mathbf{S}_d$ en un point où son gradient s'annule. Calculer une expression explicite pour $(\hat{\mu}, \hat{\Sigma})$ en fonction de $x_1, x_2, \ldots, x_n$. Vous pouvez utiliser les identités de calcul différentiel matriciel suivantes sans les démontrer :

$$\frac{\partial u^T M^{-1} v}{\partial M} = -(M^{-1})^T uv^T (M^{-1})^T,$$

$$\frac{\partial \det(M)}{\partial M} = \det(M)(M^{-1})^T,$$

où M est une matrice inversible et u et v sont des matrices colonnes de dimension compatible avec M (par exemple si M est de taille n par n, u et v sont de taille n par 1).

Solution : Commencons par calculer les gradients par rapport à μ et par rapport à Σ.

$$\begin{aligned}\nabla_\mu \log(\ell(\mu,\Sigma))) &= 0 + 0 + \nabla_\mu\left(-\sum_{i=1}^n \frac{1}{2}(x_i-\mu)\Sigma^{-1}(x_i-\mu)\right)\\ &= -\sum_{i=1}^n \frac{1}{2}\nabla_\mu((x_i-\mu)^T\Sigma^{-1}(x_i-\mu)).\end{aligned}$$

Or pour Σ symmétrique, $(x_i-\mu)^T\Sigma^{-1}(x_i-\mu) = x_i^T\Sigma^{-1}x_i - 2\mu^T\Sigma^{-1}x_i + \mu^T\Sigma^{-1}\mu$ et $\nabla_\mu(x_i^T\Sigma^{-1}\mu) = \Sigma^{-1}x_i$ et $\nabla_\mu(\mu^T\Sigma^{-1}\mu) = 2\Sigma^{-1}\mu$.

Donc :

$$\begin{aligned}\nabla_\mu \log(\ell(\mu,\Sigma))) &= -\sum_{i=1}^n \frac{1}{2}(-2\Sigma^{-1}x_i + 2\Sigma^{-1}\mu)\\ &= \Sigma^{-1}\left(\sum_{i=1}^n (x_i-\mu)\right).\end{aligned}$$

Pour le gradient par rapport à Σ, en utilisant la règle de dérivation pour les fonctions composées de plusieurs variables (par le biais des matrices jacobiennes) et les identités de calcul différentiel données ci-dessus (et la symmétrie de l'inverse d'une matrice symmétrique inversible) on obtient :

$$\begin{aligned}\nabla_\Sigma \log(\ell(\mu,\Sigma))) &= 0 - \frac{n}{2}J_{\log}(|\Sigma|)J_{|.|}(\Sigma) + \sum_{i=1}^n \frac{1}{2}\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T\Sigma^{-1}\\ &= -\frac{n}{2}\frac{1}{|\Sigma|}|\Sigma|\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}\left(\sum_{i=1}^n (x_i-\mu)(x_i-\mu)^T\right)\Sigma^{-1}\\ &= \frac{1}{2}\Sigma^{-1}\left(\left(\sum_{i=1}^n (x_i-\mu)(x_i-\mu)^T\right)\Sigma^{-1} - nI\right).\end{aligned}$$

Solution : (Continuée)

On cherche à présent $(\hat{\mu},\hat{\Sigma}) \in \mathbf{R}^d \times \mathbf{S}_d$ qui annulent ces gradients. En utilisant le fait que toute matrice de $\mathbf{S}_d$ est inversible, on obtient que $\nabla_\mu \log(\ell(\hat{\mu},\hat{\Sigma}))) = 0$ implique $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$ et $\nabla_\Sigma \log(\ell(\hat{\mu},\hat{\Sigma}))) = 0$ implique $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n (x_i-\hat{\mu})(x_i-\hat{\mu})^T$. On vérifie bien que, réciproquement, ce choix de $\hat{\mu}$ et $\hat{\Sigma}$ annule les gradients et que $(\hat{\mu},\hat{\Sigma}) \in \mathbf{R}^d \times \mathbf{S}_d$. Au final, on a obtenu que l'estimateur du maximum de vraisemblance pour la moyenne d'une loi Gaussienne multivariée est simplement la moyenne empirique usuelle et celui pour la covariance est simplement la covariance empirique usuelle (au moins dans le cas où cette dernière est inversible).

4. Montrer que le couple $(\hat{\mu}, \hat{\Sigma})$ ainsi obtenu forme une statistique suffisante pour le couple de paramètres (μ^*, Σ^*).

Solution : On va montrer qu'on peut écrire $p(x_1, \ldots, x_n; \mu^*, \Sigma^*)$ avec une expression qui ne dépend de $x_1, \ldots, x_n$ que par le biais de $\hat{\mu}$ et $\hat{\Sigma}$.

Pour simplifier la notation on note simplement (μ, Σ) pour (μ^*, Σ^*) dans la suite.

On a :

$$p(x_1, \ldots, x_n; \mu, \Sigma) = \frac{1}{(2\pi)^{nd/2}(\sqrt{|\Sigma|})^n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)\right).$$

Réécrivons la seule partie où les x_i apparaissent en fonction de $\hat{\mu}$ et $\hat{\Sigma}$:

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu) &= \sum_{i=1}^{n}\left((x_i - \hat{\mu})^T \Sigma^{-1}(x_i - \hat{\mu}) - \hat{\mu}^T \Sigma^{-1}\hat{\mu} + \mu^T \Sigma^{-1}\mu + 2(\hat{\mu}^T \Sigma^{-1} x_i - \mu^T \Sigma^{-1} x_i)\right) \\ &= n(\mu^T \Sigma^{-1}\mu + \hat{\mu}^T \Sigma^{-1}\hat{\mu} - 2\mu^T \Sigma^{-1}\hat{\mu}) + \sum_{i=1}^{n}(x_i - \hat{\mu})^T \Sigma^{-1}(x_i - \hat{\mu})\end{aligned}$$

Et :

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \hat{\mu})^T \Sigma^{-1}(x_i - \hat{\mu}) &= \mathbf{Tr}\left(\sum_{i=1}^{n}(x_i - \hat{\mu})^T \Sigma^{-1}(x_i - \hat{\mu})\right) \\ &= \sum_{i=1}^{n}\mathbf{Tr}\left((x_i - \hat{\mu})^T \Sigma^{-1}(x_i - \hat{\mu})\right) \\ &= \sum_{i=1}^{n}\mathbf{Tr}\left((x_i - \hat{\mu})(x_i - \hat{\mu})^T \Sigma^{-1}\right) \\ &= \mathbf{Tr}\left(\sum_{i=1}^{n}(x_i - \hat{\mu})(x_i - \hat{\mu})^T \Sigma^{-1}\right) \\ &= \mathbf{Tr}\left(n\hat{\Sigma}\Sigma^{-1}\right) = n\mathbf{Tr}\left(\hat{\Sigma}\Sigma^{-1}\right).\end{aligned}$$

Exercice 3 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 :

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

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.

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 4 Statistique asymptotique. ($\star$)

1. Montrer que la moyenne empirique est un estimateur fortement consistant de μ pour un échantillon $X_1, X_2, \dots, X_n$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$, c'est à dire que la moyenne empirique tend presque sûrement vers μ.
2. Montrer que l'estimateur usuel de la variance est fortement consistant.
3. Donnez une expression pour la distribution asymptotique de la norme L2 de la moyenne empirique de vecteurs aléatoires $X_1, X_2, \dots, X_n$, i.i.d. de loi P. On suppose que les moments d'ordre 1 et 2 de P existent.

Exercice 5 Intervalles de confiance. ($\star\star$)

On considère des variables aléatoires rélles X_1, X_2,…, X_n i.i.d. de loi P et de variance :

$$\mathrm{Var}(P) = \sigma^2 < +\infty.$$

On suppose de plus que X_1, X_2,…, X_n prennent leur valeurs dans l'intervalle $[0,1]$ et que la fonction de répartition de P est continue.

On définit un intervalle de confiance asymptotique pour $\mathbf{E}(P)$ de niveau $1-\alpha$, pour $\alpha \in]0;1[$, comme une région (i.e. un sous-ensemble) R_n de la droite réelle telle que :

$$\lim_{n \to +\infty} p\left(\mathbf{E}(P) \in R_n\right) \geq 1 - \alpha.$$

1. En vous appuyant sur le théorème de la limite centrale, donnez l'expression d'un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour $\mathbf{E}(P)$ utilisant l'estimateur de la moyenne empirique $\hat{\mu} := \frac{1}{n}\sum_{i=1}^{n} X_i$ et prouvez que l'intervalle proposé est effectivement asymptotiquement de niveau $1 - \alpha$.

Solution : X_1, X_2, …, X_n vérifient les conditions d'application du théorème de la limite centrale, qui donne :

$$\sqrt{n}(\hat{\mu}(X_1, X_2, \ldots, X_n) - \mathbf{E}(P)) \to_d \mathcal{N}(0, \sigma^2). \tag{1}$$

Ce résultat suggère de considérer un intervalle de confiance de la forme :

$$R_n := \left[\hat{\mu} - \frac{\sigma q(1-\alpha/2)}{\sqrt{n}}; \hat{\mu} - \frac{\sigma q(\alpha/2)}{\sqrt{n}}\right],$$

où pour tout $a \in]0;1[$, $q(a)$ est le quantile d'ordre a de $\mathcal{N}(0,1)$. On a, en effet :

$$\begin{aligned}\mathbf{E}(P) \in R_n \Leftrightarrow -\mathbf{E}(P) &\in \left[-\hat{\mu} + \frac{\sigma q(\alpha/2)}{\sqrt{n}}; -\hat{\mu} + \frac{\sigma q(1-\alpha/2)}{\sqrt{n}}\right] \\ &\Leftrightarrow \sqrt{n}(\hat{\mu} - \mathbf{E}(P)) \in [\sigma q(\alpha/2); \sigma q(1-\alpha/2)].\end{aligned}$$

et :

$$p_{X \sim \mathcal{N}(0,\sigma^2)}\left(X \in]\sigma q(\alpha/2); \sigma q(1-\alpha/2)[\right) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

Cependant, pour obtenir un intervalle de confiance calculable directement à partir de X_1, …, X_n, il nous faut un estimateur pour σ. Prenons par exemple :

$$\hat{s}(X_1, \ldots, X_n) := \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2},$$

et considérons un intervalle de confiance de la forme :

$$R_n := \left[\hat{\mu} - \frac{\hat{s} q(1-\alpha/2)}{\sqrt{n}}; \hat{\mu} - \frac{\hat{s} q(\alpha/2)}{\sqrt{n}}\right].$$

Il reste à prouver que :

$$\lim_{n \to +\infty} p\left(\mathbf{E}(P) \in R_n\right) \geq 1 - \alpha.$$

Commencons par prouver que $\hat{s}$ converge bien en probabilité vers σ. Considérons les variables aléatoires $Y_i := X_i^2$. Comme les variables Y_i sont i.i.d., $\mathbf{E}(Y_i) = \mathbf{E}(P)^2 + \sigma^2 < +\infty$ et $Y_i = |Y_i|$, la loi forte des grands nombres s'applique, donnant $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_{a.s.} \mathbf{E}(P)^2 + \sigma^2$ et donc $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_p \mathbf{E}(P)^2 + \sigma^2$.

Solution : (Continuée)

D'autre part la loi forte des grand nombres appliquée aux variables X_i suivie du théorème sur les transformations continues appliqué au cas de la fonction $x \mapsto x^2$, donne $\hat{\mu}^2 \to_{a.s.} \mathbf{E}(P)^2$ et donc $\hat{\mu}^2 \to_p \mathbf{E}(P)^2$.

De plus, on voit immédiatement en revenant à la définition de la convergence en probabilité que $Z_n \to_p Z$ et $T_n \to_p T$ implique que $Z_n + T_n \to_p Z + T$ pour tout choix de Z_n, Z, T_n, T. On en conclut que $\hat{s}^2 = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \hat{\mu}^2 \to_p \mathbf{E}(P)^2 + \sigma^2 - \mathbf{E}(P)^2 = \sigma^2$.

On applique encore une fois le théorème sur les transformations continues, cette fois pour le cas la fonction $x \mapsto \sqrt{x}$ qui est continue sur le support de $\hat{s}$ (i.e. $\mathbf{R}^+$), pour finalement obtenir :

$$\hat{s} \to_p \sqrt{\sigma^2} = \sigma \tag{2}$$

Notons F_n la fonction de répartition de $\frac{\sqrt{n}(\hat{\mu}-\mathbf{E}(P))}{\hat{s}}$ et $F = q^{-1}$ la fonction de répartition de $\mathcal{N}(0,1)$.

On a, d'une part :

$$\mathbf{E}(P) \in R_n \Leftrightarrow \frac{\sqrt{n}(\hat{\mu} - \mathbf{E}(P))}{\hat{s}} \in [q(\alpha/2); q(1-\alpha/2)],$$

et donc :

$$p(\mathbf{E}(P) \in R_n) = F_n(q(1-\alpha/2)) - F_n(q(\alpha/2)). \tag{3}$$

Et, d'autre part, en appliquant le théorème de Slutsky sur la base des résultats (1) et (2), on a :

$$\frac{\sqrt{n}(\hat{\mu} - \mathbf{E}(P))}{\hat{s}} \to_d \mathcal{N}(0,1). \tag{4}$$

La continuité de F_n se déduit de la continuité de la fonction de répartition de P. Le résultat (4) donne alors que pour tout $x \in \mathbf{R}$:

$$\lim_{n\to+\infty} F_n(x) = F(x).$$

Appliqué à (3), cela donne que :

$$\lim_{n\to+\infty} p(\mathbf{E}(P) \in R_n) = F(q(1-\alpha/2)) - F(q(\alpha/2)) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

Donc R_n est un intervalle de confiance asymptotique pour $\mathbf{E}(P)$ de niveau $1-\alpha$.

2. En vous appuyant sur l'inégalité de concentration de Hoeffding (`https://fr.wikipedia.org/wiki/In%C3%A9galit%C3%A9_de_Hoeffding`), donnez l'expression d'un intervalle de confiance (non-asymptotique) de niveau $1-\alpha$ pour $\mathbf{E}(P)$ (toujours sur la base de $\hat{\mu}$) et prouvez que l'intervalle proposé est de niveau $1-\alpha$.

Solution : $X_1, X_2, \ldots, X_n$ vérifient les conditions d'application de l'inégalité de Hoeffding, qui donne pour tout $t > 0$:

$$p\left(\left|\sum_{i=1}^{n} X_i - \mathbf{E}\left(\sum_{i=1}^{n} X_i\right)\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{n}\right).$$

Soit $t > 0$. En divisant par n et en utilisant la linéarité de l'intégrale, on obtient que : $\left|\sum_{i=1}^{n} X_i - \mathbf{E}\left(\sum_{i=1}^{n} X_i\right)\right| \geq t$ si et seulement si $|\hat{\mu} - \mathbf{E}(P)| \geq \frac{t}{n}$.
On en déduit que :

$$p\left(|\hat{\mu} - \mathbf{E}(P)| \geq \frac{t}{n}\right) \leq 2\exp\left(-\frac{2t^2}{n}\right).$$

Pour obtenir un intervalle de confiance de niveau $1-\alpha$, il nous faut identifier une région $\mathcal{R}$ de la droite réelle telle que $p(\mathbf{E}(P) \in \mathcal{R}) \geq 1-\alpha$.
En prenant $\alpha = 2\exp\left(-\frac{2t^2}{n}\right)$, c'est à dire $t = \sqrt{-\frac{n}{2}\log\left(\frac{\alpha}{2}\right)}$, on voit que la région :

$$\mathcal{R} = \left[\hat{\mu} - \sqrt{\frac{-\log\left(\frac{\alpha}{2}\right)}{2n}}; \hat{\mu} + \sqrt{\frac{-\log\left(\frac{\alpha}{2}\right)}{2n}}\right]$$

convient.

3. Comparez la largeur des intervalles obtenus par les deux méthodes et commentez.

Solution : Les deux intervalles de confiance que l'on a proposé sont centrés sur $\hat{\mu}$ et symétriques de part et d'autre de $\hat{\mu}$, ils diffèrent donc uniquement par leur largeur.
En utilisant la symmétrie de la loi normale, on obtient que la largeur de l'intervalle de confiance obtenu avec le théorème de la limite centrale est de :

$$\ell_{TCL} := \frac{\hat{s}(q(1-\alpha/2) - q(\alpha/2))}{\sqrt{n}} = \frac{2\hat{s}q(1-\alpha/2)}{\sqrt{n}},$$

celle de l'intervalle de confiance obtenu avec l'inégalité de Hoeffding est de :

$$\ell_H := \sqrt{\frac{-2\log\left(\frac{\alpha}{2}\right)}{n}}.$$

Solution : (Continuée)

On voit que la dépendance à la taille de l'échantillon est en $O(1/\sqrt{n})$ dans les deux cas, ce qui correspond à la vitesse de convergence statistique usuelle en l'absence d'hypothèses de régularité particulières.

Une première différence importante est que la largeur de l'intervalle de confiance obtenu avec le théorème de la limite centrale est proportionnelle à l'écart-type (estimé) de P (elle s'*adapte* à la variance de P), là où l'approche basé sur l'inégalité de Hoeffding n'utilise que le fait que les variables soient bornées. L'écart-type peut en général être estimé de manière fiable pour des n raisonnables et l'intervalle obtenu avec le théorème de la limite centrale devrait donc typiquement être plus précis (et beaucoup moins conservateur), dans le cas où la distribution P est concentrée sur une petite portion de l'intervalle $[0; 1]$. Notons cependant qu'il est possible de développer des inégalités de concentration prenant en compte la variance de la distribution (moyennant certaines hypothèses de régularité) et que pour aller plus loin, il serait intéressant de comparer la largeur d'intervalles de confiances obtenus sur la base de ces inégalités à ℓ_{TCL}.

Pour comparer les deux approches de manières équitable au-delà de cette première différence, on peut considérer le cas où la variance est maximale pour des variables aléatoires sur $[0; 1]$. La variance dans ce cas est de $1/4$ (cf. l'inégalité de Popiviciu, saturée par une distribution discrète prenant les valeurs 0 et 1 avec probabilité $1/2$), conduisant à un écart-type de $1/2$. α est typiquement relativement petit et il paraît donc naturel de considérer l'équivalent asymptotique $q(1-x) \sim_{x \to 0^+} \sqrt{-2\log(x)}$, qui nous donne :

$$\ell_{TCL} \approx \sqrt{\frac{-2\log(\frac{\alpha}{2})}{n}} = \ell_H$$

On voit que dans le cas d'une distribution de variance maximale, les deux approches donnent des intervalles de confiance de largeur asymptotiquement équivalentes quand la précision $1-\alpha$ de l'intervalle de confiance tend vers 1 (attention, notez que le terme *asymptotique* ici réfère à α et pas au cadre statistique considéré, n étant fixe). Cette équivalence asymptotique ne doit pas masquer cependant des différences non négligeables pour des valeurs de α utilisées en pratique et une analyse plus poussées pourrait être utilisée pour montrer que l'intervalle obtenue par l'inégalité de Hoeffding est toujours le plus conservateur. Par exemple, pour des intervalles de confiance à 95% et à 99%—même dans le cas d'une distribution de variance maximale—l'intervalle obtenu avec l'inégalité de Hoeffding est respectivement environ 1.4 et 1.3 fois plus large que l'intervalle obtenu avec le théorème central limite.