Aix*Marseille université

Année 2025-26

MASTER 2 INFORMATIQUE

PARCOURS IAAA, UE MIA

Feuille d'exercices numéro 4

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

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.
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)

5. Exprimez l'espérance de $\hat{V}$ comme une fonction de l'espérance et des moments centrés des variables $X_1, X_2, \ldots, X_n$ (prouvez votre résultat)
6. Exprimez la variance de $\hat{V}$ comme une fonction de l'espérance et des moments centrés des variables $X_1, X_2, \ldots, X_n$ (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.
2. Calculer le biais de l'estimateur de la variance suivant pour un échantillon i.i.d. : $\hat{V} := \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$
3. Proposez un estimateur non-biaisé de la variance pour un échantillon i.i.d.
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 $\hat{\delta}$?

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 $\sigma_2^2 = \mathrm{Var}_{(x_1,x_2) \sim P \otimes P}[d(x_1, x_2)]$.

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

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

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 $\hat{\delta} \to_p \delta$, c'est à dire que pour tout $\epsilon > 0$:
$$\lim_{n \to +\infty} p(|\hat{\delta}(x_1, \ldots x_n) - \delta| \geq \epsilon) = 0.$$

Exercice 6 On considère des variables aléatoires réelles $X_1, X_2, \ldots, X_n$ 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 :
$$\hat{\mu}(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

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 $X_1, X_2, \ldots X_n$ identiquement distribuées de loi P, *mais pas forcément indépendantes.* On note $X := [X_1|X_2|\ldots|X_n]$ 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 $c_{i,i}$?
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.
3. Pour un σ fixé, quel choix de Σ maximise la variance de $\hat{\mu}$? Quel choix minimise sa variance ?

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

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).