Aix*Marseille université

Année 2025-26

MASTER 2 INFORMATIQUE

PARCOURS IAAA, UE MIA

Feuille d'exercices numéro 3

Probabilités et maximum de vraisemblance.

1 Calcul probabiliste

Exercice 1 Propriétés élémentaires. ($\star$)

1. Prouvez que si A et B sont des événements indépendants, $P(A|B) = P(A)$.
2. On considère une fonction f de $\mathbf{R}$ vers $\mathbf{R}$. Quelles conditions doit-elle vérifier pour être une fonction de répartition valide ? Une densité de probabilité valide ?
3. Représentez graphiquement la fonction de masse, la densité de probabilité et la fonction de répartition pour une variable aléatoire X uniforme sur $[0, 1]$.
4. Même question pour X valant 1 avec probabilité .5, 2 avec probabilité .3 et 4 avec probabilité .2.
5. Un coffre A contient 100 pièces d'or. Un coffre B contient 60 pièces d'or et 40 pièces d'argent. Vous choisissez un coffre aléatoirement selon une loi uniforme et tirez une pièce aléatoirement selon une loi uniforme dans ce coffre. Si la pièce est en or, quelle est la probabilité que vous ayez choisi le coffre A ?

Exercice 2 Calcul probabiliste ($\star$)

On considère à présent une variable aléatoire binaire Z et deux variables aléatoires réelles X et Y telles que $P(X, Y, Z) = P(X|Z)P(Y|Z)P(Z)$ avec

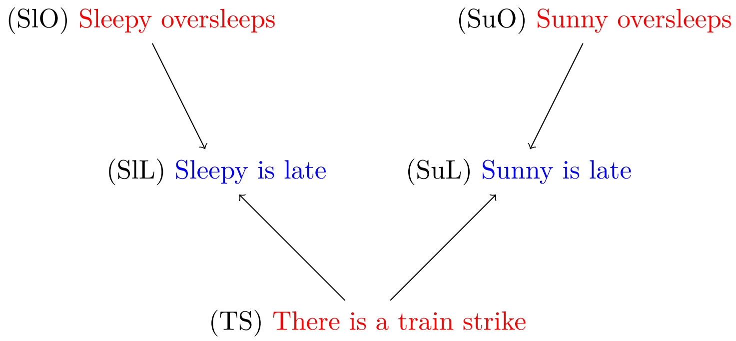
1. $P(Z = 1) = 2/3$
2. $P(X|Z = 0) = \mathcal{N}(X; 0, 1)$
3. $P(X|Z = 1) = \mathcal{N}(X; 1, 1)$
4. $P(Y|Z = 0) = \mathcal{N}(Y; 0, 1)$
5. $P(Y|Z = 1) = \mathcal{N}(Y; -1, 1)$

où

$$\mathcal{N}(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

est la densité d'une loi normale de moyenne μ et de variance σ^2. On suppose qu'on observe $Y = y_0$. Calculez $P(X|Y = y_0)$.

Exercice 3 Calcul probabiliste. (⋆⋆)



Considérons le modèle graphique représenté ci-dessus. SlO, SuO, SlL, SuL and TS sont des variables aléatoire binaires prenant leurs valeurs dans $\{0, 1\}$. Dans cet exercice nous allons essayer de déterminer ce qui peut être inféré sur les variables latentes (en rouge) à partir de l'observation des variables observées (en bleu).

Nous faisons l'hypothèse que si au moins un des deux évènements "Sleepy oversleeps" et "There is a train strike" a lieu, alors 'Sleepy is late' a lieu également (avec probabilité 1). De façon similaire, si au moins un des deux évènements 'Sunny oversleeps' et 'There is a train strike' a lieu, alors 'Sunny is late' a lieu également (avec probabilité 1). Nous pouvons l'écrire plus formellement, de la manière suivante :

$$P(SlL = 1|SlO = a, TS = b) = a \vee b,$$

et :

$$P(SuL = 1|SuO = a, TS = b) = a \vee b,$$

pour tout a, b dans $\{0, 1\}$. Le symbole $\vee$ représente le connecteur logique *ou* (inclusif) de $\{0, 1\}$ vers $\{0, 1\}$.

On note $l = P(SlO = 1)$, $u = P(SuO = 1)$ et $t = P(TS = 1)$.

1. Donner la factorisation de $P(SlL, SuL, SlO, SuO, TS)$ d'après le modèle graphique représenté ci-dessus.
2. La distribution de probabilité $P(SlL, SuL, SlO, SuO, TS)$ est-elle entièrement déterminée si les valeurs de l, u et t sont données ?
3. Calculer $P(TS = 1|SlL = 1)$ en fonction de l, u et t.
4. Calculer $P(SlO = 1|SlL = 1)$ en fonction de l, u et t.
5. Calculer $P(TS = 1|SlL = 1, SuL = 1)$ en fonction de l, u et t.

6. Calculer $P(SlO = 1|SlL = 1, SuL = 1)$ en fonction de l, u et t.

7. Supposer à présent que $l = 0.5$, $t = 0.1$ et que l'évenement 'Sleepy is late' a lieu. Quel évènement est alors le plus probable : 'There is a train strike' ou 'Sleepy overslept' ?

8. Même question si on suppose en plus que $u = 0.01$ et que l'évènement 'Sunny is late' est également observé.

9. Que se passe-t-il si on prend $l = 0.5$, $t = 0.1$ et $u = 0.2$?

Exercice 4 Calcul probabiliste ($\star$)

On considère quatre variables aléatoires X_1, X_2, Z_1, Z_2, dont la loi jointe s'écrit $p(X_1, X_2, Z_1, Z_2) = p(X_1|Z_1)p(X_2|Z_2)p(Z_2|Z_1)p(Z_1)$, avec :

— Z_1 binaire, $p(Z_1 = 1) = p$.

— Z_2 binaire, $p(Z_2 = 1|Z_1 = 0) = q_1$ et $p(Z_2 = 1|Z_1 = 1) = q_2$

— $p(X_1 = x|Z_1 = 0) \sim \mathcal{N}(x; 0, 1)$

— $p(X_1 = x|Z_1 = 1) \sim \mathcal{N}(x; 1, 1)$

— $p(X_2 = x|Z_2 = 0) \sim \mathcal{N}(x; 0, 1)$

— $p(X_2 = x|Z_2 = 1) \sim \mathcal{N}(x; -1, 1)$

où $\mathcal{N}(x; \mu, \sigma^2)$ donne la densité d'une loi gaussienne de moyenne μ et de variance σ^2 en x.

1. On suppose qu'on observe $X_1 = x_1$. Donnez une expression pour $p(X_2|X_1 = x_1)$ en fonction de x_1, p, q_1 et q_2.

2. On suppose à présent qu'on observe $X_1 = 0$ et $X_2 = 0$. Donnez une expression pour $p(X_1 = 0, X_2 = 0)$ en fonction de p, q_1 et q_2, calculez son gradient par rapport à p, q_1 et q_2 et donnez les valeurs de p, q_1 et q_2 pour lesquels ce gradient s'annule.

3. Que est l'intérêt de ce calcul ?

Exercice 5 Calcul probabiliste ($\star$)

On considère quatre variables aléatoires X_1, X_2, Z_1, Z_2, dont la loi jointe s'écrit $p(X_1, X_2, Z_1, Z_2) = p(X_1|Z_1)p(X_2|Z_2)p(Z_1)p(Z_2)$, avec pour $i \in 1, 2$:

— $p(Z_i = 0) = p$, $p(Z_i = 1) = q$, $p(Z_i = 2) = 1 - p - q$

— $p(X_i = x|Z_i = 0) \sim \mathcal{N}(x; 0, 1)$

— $p(X_i = x|Z_i = 1) \sim \mathcal{N}(x; 1, 1)$

— $p(X_i = x|Z_i = 2) \sim \mathcal{N}(x; -1, 1)$

où $\mathcal{N}(x; \mu, \sigma^2)$ donne la densité d'une loi gaussienne de moyenne μ et de variance σ^2 en x.

1. On suppose qu'on observe $X_1 = x_1$. Donnez une expression pour $p(Z_1|X_1 = x_1)$ en fonction de x_1, p, et q.
2. On suppose qu'on observe $X_2 = x_2$. Donnez une expression pour $p(Z_1|X_2 = x_2)$ en fonction de x_2, p, et q.
3. On suppose à présent qu'on observe $X_1 = 0$ et $X_2 = 0$. Donnez une expression pour $p(X_1 = 0, X_2 = 0)$ en fonction de p et q.
4. Trouvez les valeurs de p et q qui maximisent $p(X_1 = 0, X_2 = 0)$ (justifiez votre réponse).
5. Que est l'intérêt de ce calcul ?

2 Maximum de vraisemblance

Exercice 6 Estimation par maximum de vraisemblance des paramètres d'une loi Gaussienne multivariée. ($\star$)

Supposons qu'on observe un échantillon i.i.d. $x_1, x_2, \dots, x_n \in (\mathbf{R}^d)^n$ 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 $x_1, x_2, \dots, x_n$.

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

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

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

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

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

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

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).

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