

Scribe séance 4

Noms

September 26, 2024

1 Le Problème Primal

Nous considérons le problème d'optimisation suivant :

$$\begin{array}{ll}\text{Minimiser} & f(x) \\ \text{Sous les contraintes} & g_j(x) \leq 0, \quad j = 1, \dots, r \\ & x \in E \subseteq \mathbb{R}^n\end{array}$$

où :

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est la fonction objectif.
- $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ sont les fonctions de contrainte.

La valeur optimale du problème primal est :

$$f^* = \inf_{x \in E, g_j(x) \leq 0} f(x)$$

2 Le Lagrangien Associé

Nous définissons la fonction lagrangienne :

$$L(x, \mu) = f(x) + \sum_{j=1}^r \mu_j g_j(x)$$

où $\mu = (\mu_1, \dots, \mu_r)$ sont les multiplicateurs de Lagrange, avec $\mu_j \geq 0$.

3 La Fonction Duale

La fonction duale $q(\mu)$ est définie par :

$$q(\mu) = \inf_{x \in E} L(x, \mu)$$

Cette fonction représente la valeur minimale du lagrangien par rapport à x pour un μ donné.

4 Le Problème Dual

Le problème dual est :

$$\begin{array}{ll}\text{Maximiser} & q(\mu) \\ \text{Sous les contraintes} & \mu_j \geq 0, \quad j = 1, \dots, r\end{array}$$

5 La Valeur Optimale du Problème Dual

La valeur optimale du problème dual est :

$$g^* = \sup_{\mu \geq 0} q(\mu)$$

5.1 Résultat clé

Dualité faible : Pour tout x admissible et μ admissible, on a $q(\mu) \leq f(x)$. Par conséquent :

$$g^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in E, g_j(x) \leq 0} f(x) = f^*$$

Cela implique que la valeur optimale du problème dual est une borne inférieure de la valeur optimale du problème primal.

6 Propriétés de Convexité

- Le domaine D_q de la fonction duale q est convexe.
- La fonction duale q est concave sur D_q .

Cela s'explique par le fait que q est l'infimum de fonctions affines (linéaires) en μ , ce qui préserve les propriétés de convexité.

7 Dualité Forte et Qualification des Contraintes

Sous certaines conditions, appelées qualifications des contraintes, la dualité forte s'applique, c'est-à-dire :

$$g^* = f^*$$

7.1 Exemple

Si f est convexe et les fonctions de contrainte g_j sont linéaires (affines), alors la dualité forte s'applique et il n'y a pas d'écart dual.

8 Conditions d'Optimalité (Conditions de Karush-Kuhn-Tucker)

Proposition (Conditions KKT) : Supposons que f et g_j sont différentiables. Les points x^* et μ^* sont optimaux pour le problème primal (respectivement dual) si et seulement s'ils satisfont :

1. Faisabilité Primal :

$$g_j(x^*) \leq 0, \quad \text{pour tout } j = 1, \dots, r$$

2. Faisabilité Duale :

$$\mu_j^* \geq 0, \quad \text{pour tout } j = 1, \dots, r$$

3. Complémentarité :

$$\mu_j^* \cdot g_j(x^*) = 0, \quad \text{pour tout } j = 1, \dots, r$$

4. Stationnarité :

$$\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) = 0$$

Complémentarité :

$$\mu_j^* \cdot g_j(x^*) = 0, \quad \text{pour tout } j = 1, \dots, r$$

Stationnarité :

$$\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) = 0$$

Ces conditions sont nécessaires pour l'optimalité, et si f et g_j sont convexes, elles sont également suffisantes.

9 Interprétation

- **Faisabilité Primal** : x^* satisfait les contraintes du problème initial.
- **Faisabilité Duale** : Les multiplicateurs μ^* sont non négatifs.
- **Complémentarité** : Si une contrainte n'est pas active ($g_j(x^*) < 0$), alors le multiplicateur correspondant est nul ($\mu_j^* = 0$).
- **Stationnarité** : Le gradient du lagrangien par rapport à x est nul en x^* , indiquant un point critique.

10 Problèmes Généraux d'Estimation

10.1 Modèle Statistique

- **Modèle** : Une famille de distributions de probabilité \mathcal{P} .
- **Observations** : $O \sim P$, où $P \in \mathcal{P}$.
- **But** : Estimer une fonctionnelle $f(P)$ à partir des observations O , c'est-à-dire trouver un estimateur $\hat{f}(O)$ tel que $\hat{f}(O) \approx f(P)$.

10.2 Fonctionnelle et Estimateur

- **Fonctionnelle** : $f : \mathcal{P} \rightarrow E$, où E est un espace cible (par exemple, \mathbb{R} ou \mathbb{R}^d).
- **Estimateur** : $\hat{f} : O \rightarrow E$.

11 Deux Questions Fondamentales en Estimation

1. Trouver un estimateur efficace sur le plan calculatoire : Développer un estimateur \hat{f} qui est computationnellement réalisable et pratique à utiliser.
2. Évaluer la qualité statistique de l'estimateur : Analyser dans quelle mesure $\hat{f}(O)$ approche $f(P)$, en termes de biais, variance et risque.

Pour les cas où nous estimons une distribution, cela implique :

- Calculs de probabilités (par exemple, estimation du maximum de vraisemblance).
- Contrôle des moments ou des queues de distribution (par exemple, évaluation du biais, de la variance, du risque attendu).

11.1 Exemple A : Estimation du Maximum de Vraisemblance

- **Modèle Statistique** : Une famille paramétrique de distributions $\mathcal{P} = \{P_w \mid w \in W\}$, où W est l'ensemble des paramètres.
- **Observations** : $O \sim P_w$ pour un certain $w \in W$.
- **Fonctionnelle à estimer** : $f(P_w) = w$, c'est-à-dire, nous souhaitons estimer le paramètre w lui-même.
- **Densité** : p_w est la densité de probabilité associée à P_w (par rapport à une mesure fixe).

Estimateur du Maximum de Vraisemblance :

$$\hat{f}_H(O) = \arg \max_{w \in W} p_w(O)$$

Nous choisissons le paramètre w qui maximise la vraisemblance des observations O .

Évaluation de l'estimateur :

- **Biais** : Différence entre la valeur moyenne de l'estimateur et le véritable paramètre.
- **Variance** : Variabilité de l'estimateur autour de sa moyenne.
- **Risque** : Mesure globale de l'erreur de l'estimateur, souvent définie via une fonction de perte.

11.2 Exemple B : Qualité Statistique d'un Estimateur

Biais :

$$b_P(\hat{f}) = \mathbb{E}_{O \sim P}[\hat{f}(O)] - f(P)$$

Amplitude du biais :

$$\|b_P(\hat{f})\|^2$$

11.3 Variance

$$\text{Var}_P(\hat{f}) = \mathbb{E}_{O \sim P} \left[\left\| \hat{f}(O) - \mathbb{E}_{O \sim P}[\hat{f}(O)] \right\|^2 \right]$$

11.4 Risque

$$R_P(\hat{f}) = \mathbb{E}_{O \sim P}[\ell(f(P), \hat{f}(O))]$$

- **Fonction de coût** : $\ell : x, y \mapsto \|x - y\|_2^2$

11.5 Décomposition du risque

$$R_P(\hat{f}) = \text{Var}_P(\hat{f}) + \|b_P(\hat{f})\|_2^2$$

Cette décomposition montre que le risque total est la somme de la variance et du carré du biais de l'estimateur.

12 Probabilité et Théorie de la Mesure

12.1 Espace de Probabilité

- **Univers** (Ω): Désigne l'ensemble de tous les résultats possibles liés à une expérience aléatoire.
- **Événement** (A): L'événement $A \subset \Omega$ est un sous-ensemble de l'univers, ce qui le définit comme un ensemble de résultats potentiels.
- **Ensemble des événements (Tribu \mathcal{F})** : Représente une collection d'ensembles d'événements $\mathcal{F} \subset \mathcal{P}(\Omega)$ qui forme un espace de probabilités. Cette collection doit être stable sous les unions dénombrables et par compléments.
- **Mesure de probabilité** : La mesure de probabilité est une fonction qui attribue une valeur numérique à un événement, représentant la chance que cet événement se produise. La probabilité d'un événement varie de 0 (l'événement ne peut pas se produire) à 1 (l'événement se produira sûrement).

On peut donc l'exprimer par $P : \mathcal{F} \rightarrow \mathbb{R}$, vérifiant :

- $\forall A \in \mathcal{F}, P(A) \geq 0$
- $P(\Omega) = 1$ (La somme des probabilités de tous les événements possibles dans un espace d'échantillonnage est égale à 1.)
- Pour toute suite dénombrable d'événements disjoints $(A_i)_{i \geq 1} \subseteq \mathcal{F}$:

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i)$$

12.2 Variable Aléatoire

Une variable aléatoire X est une fonction qui associe à chaque résultat d'une expérience aléatoire (Ω) un nombre réel.

- Une fonction mesurable $X : \Omega \rightarrow E$, où (E, \mathcal{E}) est un espace mesurable.
- **Tribu sur l'espace d'arrivée**: Une collection d'événements (sous-ensembles) de E qui satisfait certaines propriétés (comme la fermeture sous les unions et les compléments).

$$\mathcal{T} \subset \mathcal{P}(E)$$

$$\forall T \in \mathcal{T}, X^{-1}(T) \in \mathcal{F}$$

- **Mesure image induite** : Pour tout $T \in \mathcal{E}$:

$$P_X(T) = P(X^{-1}(T))$$

P_X est la loi de X .

13 Probabilité Conditionnelle et Indépendance

13.1 Probabilité Conditionnelle

Si $P(B) > 0$, la probabilité d'un événement A sachant que l'événement B a eu lieu est donnée par :

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

13.2 Indépendance Mutuelle

Des variables aléatoires X_1, X_2, \dots, X_n sont mutuellement indépendantes si, pour tout A_1, A_2, \dots, A_n mesurables :

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i)$$

14 Variables Aléatoires Réelles

14.1 Fonction de Masse

La fonction de masse pour une variable aléatoire réelle X est définie par :

$$p_X(x) = P(X = x)$$

Elle donne la probabilité que la variable aléatoire X prenne la valeur x .

14.2 Fonction de Répartition

La fonction de répartition, également appelée fonction de distribution cumulative, est donnée par :

$$F_X(x) = P(X \leq x)$$

Elle représente la probabilité que la variable aléatoire X soit inférieure ou égale à une certaine valeur x .

14.3 Densité de Probabilité

La densité de probabilité décrit la distribution des probabilités sur l'ensemble des valeurs possibles de X . Pour les variables aléatoires continues, elle est calculée comme la dérivée de la fonction de répartition :

$$p_{dX}(x) = \frac{dF_X(x)}{dx}$$

La fonction de densité de probabilité doit être telle que l'intégrale sur l'ensemble des valeurs possibles soit égale à 1.

$$\int_{-\infty}^{+\infty} p_{dX}(x) dx = 1$$

14.4 Espérance

L'espérance d'une variable aléatoire X est la valeur moyenne attendue de X . Elle est donnée par :

14.5 Cas continu

$$E[X] = \int_{-\infty}^{+\infty} x p_X(x) dx = \int_{-\infty}^{+\infty} x dF_X(x)$$

14.6 Cas discret

$$E[X] = \sum_x x p_X(x)$$

15 Plusieurs Variables Aléatoires Réelles

On peut définir plusieurs variables aléatoires réelles par $\mathbf{X} = (X_1, X_2, \dots, X_N)$, un vecteur de la fonction de Ω vers \mathbb{R}^n .

$$\mathbf{X} = (X_1, X_2, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$$

15.1 Fonction de répartition (jointe)

Pour des variables aléatoires multiples (X_1, X_2, \dots, X_n) , la fonction de répartition jointe est donnée par :

$$F_X(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

Elle représente la probabilité que chaque variable X_i soit inférieure ou égale à x_i .

15.2 Fonction de masse (jointe)

La fonction de masse jointe pour des variables discrètes (X_1, X_2, \dots, X_n) est définie par :

$$p_X(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

Elle donne la probabilité conjointe que chaque variable X_i prenne la valeur x_i .

15.3 Densité de probabilité (jointe)

La densité de probabilité jointe pour (X_1, X_2, \dots, X_n) est définie par :

$$p_{d\mathbf{X}} : x_1, x_2, \dots, x_n \mapsto \frac{\partial F_{\mathbf{X}}}{\partial x_1 \partial x_2 \dots \partial x_n}(x_1, x_2, \dots, x_n)$$

15.4 Espérance (jointe)

L'espérance d'une fonction de plusieurs variables aléatoires est donnée par :

- **Cas continu :**

$$E[f(X_1, X_2, \dots, X_n)] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) p_X(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

- **Cas discret :**

$$E[f(X_1, X_2, \dots, X_n)] = \sum_{x_1} \sum_{x_2} \dots \sum_{x_n} f(x_1, x_2, \dots, x_n) p_X(x_1, x_2, \dots, x_n)$$

16 Modèles graphiques

...

17 Moments et Variance

17.1 Variance

La variance d'une variable aléatoire X mesure la dispersion de X autour de son espérance $E[X]$. Elle est définie par :

$$\text{Var}[X] = E[(X - E[X])^2]$$

17.2 Moment

Le moment d'ordre k représente la dispersion de la variable aléatoire X , ce qui permet de caractériser différentes propriétés de la distribution, comme la moyenne ($k = 1$) et la variance ($k = 2$).

$$E[X^k]$$

17.3 Moment centré

$$E[(X - E[X])^k]$$

17.4 Linéarité de l'Espérance

L'espérance est une opération linéaire. Pour une somme pondérée de variables aléatoires, on a :

$$E \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i E[X_i]$$

où a_i sont des constantes réelles.

17.5 Covariance

La covariance entre deux variables aléatoires X et Y est donnée par :

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Elle mesure la tendance conjointe de X et Y à dévier de leurs espérances respectives. La covariance est bilinéaire, ce qui implique que pour des combinaisons linéaires de variables, on a :

$$\text{Cov} \left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)$$

18 Moments Conditionnels

18.1 Espérance Conditionnelle

L'espérance conditionnelle de X sachant $Y = y$ est définie par :

$$E[X | Y = y] = \int_{-\infty}^{+\infty} x p_{X|Y}(x | y) dx$$

où $p_{X|Y}(x | y)$ est la densité conditionnelle de X donnée $Y = y$.

18.2 Variance Conditionnelle

La variance conditionnelle de X sachant $Y = y$ est :

$$\text{Var}[X | Y = y] = E[(X - E[X | Y = y])^2 | Y = y]$$

19 Loi de l'Espérance Totale

La loi de l'espérance totale permet d'exprimer l'espérance d'une variable aléatoire X en fonction de l'espérance conditionnelle par rapport à une autre variable Y :

$$E[X] = E[E[X | Y]]$$

20 Loi de la Covariance Totale

De manière similaire, la covariance totale est donnée par :

$$\text{Cov}(X, Y) = \text{Cov}(E[X | Z], E[Y | Z]) + E[\text{Cov}(X, Y | Z)]$$

21 Modèles Graphiques

Les modèles graphiques permettent de représenter les dépendances entre variables aléatoires au moyen de graphes probabilistes. Pour un ensemble de variables X_1, X_2, \dots, X_n , la probabilité conjointe peut être factorisée selon :

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i))$$

où $\text{Parents}(X_i)$ représente l'ensemble des variables dont X_i dépend directement dans le graphe.