M2 IAAA

Semestre 1

Cours de MIA du jeudi 14 septembre

14 septembre 2023

1 Moments

On note la "bilinéarité" de la variance comme :

$$V[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)$$

.

La bilinéarité de la variance nous vient enfaite de la bilinéarité de la covariance tel que :

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

2 Loi gaussienne multivariée et équivalents

On pourra citer plusieur exemple de distribution tel que la Bernouilli(p) qui vaut p si on a un succès et $1-p$ sinon ou encore la loi Poisson(λ) avec comme fonctio de masse : $e^{-\lambda}\frac{\lambda^k}{k!}$

2.1 Random vectors

Les randoms vectors sont utiles pour l'application du TCL.

Qu'est-ce qu'un random vectos ?

Soit $X_1, ..., X_n$, n variable aléatoire de loi normal. On note

$$X = [X_1, ..., X_n]^T$$

le random vecteurs

2.2 Matrice de covariance

Si on a $X \in \mathbb{R}^n$, alors la matrice de covariance sera de dimension $\Sigma \in M_{n*n}(R)$ tel que

$$\Sigma = \begin{bmatrix} Cov(X_1, X_1) & ... & Cov(X_1, X_n) \\ ... & ... & ... \\ Cov(X_n, X_1) & ... & Cov(X_n, X_n) \end{bmatrix}$$

Ces variables ne sont pas iid, il y a donc une dependance entre les covariances. On peut alors écrire Σ tel que

$$\Sigma = E[(X - E(X))(X - E(X))^T]$$

$(X - E(X))$ est de dimension $1 \times n$ et $(X - E(X))^T$ de $n \times 1$ donc on a bien Σ de dimension n

Σ est symetrique semi-définie positif et si nos variables sont indépendantes entre elle, alors Σ est une matrice diagonale avec les variances de chaque variables dans sa diagonales.

2.3 Multivariate Gaussian

Soit $X \sim \mathcal{N}(\mu, \Sigma)$, $X \in \mathbb{R}$:

$$p(X; \mu; \Sigma) = \frac{1}{det(\Sigma)^{\frac{1}{2}}}(2\pi)^{\frac{n}{2}} exp[-\frac{1}{2}(x-\mu)^T \Sigma (x-\mu)]$$

$(x - \mu)$ de dimension $1 \times n$
$(x - \mu)^T$ de dimension $n \times 1$
Σ de dimension $n \times n$
Leur multiplication est donc de taille 1×1

Si on a $n = 1$, on a affaire a une gaussienne univarié

2.4 Some Nice Properties of MV Gaussians

Les marginales et les conditionnelles d'une distribution gaussienne conjointe sont gaussiennes.

Une distribution gaussienne en d dimensions, notée $X \sim \mathcal{N}(\mu, \Sigma = \text{diag}(\sigma_1^2, .., \sigma_n^2))$, est équivalente à une collection de d gaussiennes indépendantes $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Cela se traduit par des isocontours alignés sur les axes de coordonnées.

En général, les isocontours d'une distribution gaussienne multivariée sont des ellipsoïdes n-dimensionnels avec des axes principaux dans les directions des vecteurs propres de la matrice de covariance Σ. Les longueurs relatives des axes dépendent des valeurs propres de Σ.

2.5 Visualisation des gaussiennes multivariées (MV Gaussiennes)

On suppose que $Cov(X_1, X_2) = 0$ dans un premier temps

On peut voir dans les graphiques des slides que réduires les coef diagonaux de la matrice de covariance augmente la densité autour de 0 et l'augmenter rend le cerle beaucoup plus plat.

Si l'on réduit la premier composante, ça réduit sur son axe correspondant et inversement.

Maintenant on suppose que $Cov(X_1, X_2) \neq 0$ dans un second temps.

Avoir des coefficients positifs rend la distributions non symetrique et fait une rotations autour de d'un axe et si négatif une rotation autour de l'autre axe.

2.6 Multivariate Gaussian

On a que $X \sim \mathcal{N}(\mu, \Sigma) \iff \exists \mu \in \mathbb{R}^k, A \in \mathbb{R}^{k \times l}$ tel que $X = AZ + \mu$ pour $Z_n \sim \mathcal{N}(0, 1)$, iid

La distribution conditionelle nous donne alors dans ce cas, pour $X = [X_1, X_2]^T$ de taille $[q \times 1, (N-q) \times 1]^T$, $\mu = [\mu_1, \mu_2]^T$ de taille $[q \times 1, (N-q) \times 1]^T$ et

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

de taille

$$\begin{bmatrix} q \times q & q \times N - q \\ N - q \times q & N - q \times N - q \end{bmatrix}$$

On a donc les probabilité conditionelle suivant :
$p(X_1 | X_2 = a) = \mathcal{N}(\bar{\mu}, \bar{\Sigma})$ où

$$\bar{\mu} = \mu_1 + \Sigma_{12} \Sigma_{22}^{+} (a - \mu_2)$$

et

$$\bar{\Sigma} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{+} \Sigma_{21}$$

Séance 5, partie 2 : Thomas Allouche

En intelligence artificielle, la compréhension des statistiques est essentielle pour analyser et interpréter des données complexes.

1 Notations

— X est une variable aléatoire :

$$(X_1, \ldots, X_n) \to Z_n = (X_n, Y_n)$$

— L'identifiant i fait référence à la distribution de Q.
— Le modèle statistique est représenté par l'espace R^{D+1}, où D est la dimension de l'espace.

2 Fonctions d'Estimation

L'un des objectifs principaux en statistiques est d'estimer certaines fonctions à partir de données observées. Nous nous sommes intéressés à l'estimation de la fonction :

$$f(x) = \mathbb{E}_Q[Y|X = x]$$

Cette fonction représente l'espérance conditionnelle de Y étant donné $X = x$ sous la distribution Q.

— L'estimateur $\hat{f}$ est une approximation de f basée sur un échantillon de données. Il prend un ensemble de n observations (chacune étant dans R^{d+1}) et renvoie une fonction approximative de R^d vers R.

L'importance de l'estimation réside dans la capacité à prédire ou à comprendre le comportement d'une variable en fonction d'autres variables observées.

3 Caractéristiques d'un Estimateur

Pour qu'un estimateur soit considéré comme efficace et utile, il est important qu'il combine de bonnes capacités statistiques et computationnelles. Une bonne capacité statistique signifie que l'estimateur doit être capable de fournir une estimation précise et correcte basée sur les données disponibles, tandis qu'une bonne capacité computationnelle signifie que l'estimateur doit être calculable dans un délai raisonnable.

4 Approches d'Estimation

En statistique, il existe deux cadres principaux pour l'estimation : l'approche bayésienne et l'approche fréquentiste.

4.1 Approche Bayésienne

L'approche bayésienne repose sur l'utilisation d'informations préalables combinées avec les données observées pour mettre à jour les croyances sur un phénomène. Dans cette approche, la probabilité est vue comme un degré de croyance ou une mesure subjective de l'incertitude.

4.2 Approche Fréquentiste

L'approche fréquentiste, quant à elle, s'appuie sur des procédures qui ont des propriétés dans des répétitions d'une expérience. Les conclusions sont basées sur des échantillons de données, et les probabilités sont interprétées en termes de fréquences.

5 Qualité d'un Estimateur

5.1 Biais

Le biais de l'estimateur $\hat{f}$ est défini comme la différence entre l'espérance de l'estimateur et la véritable fonction f. Il est défini par :

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

5.2 Variance

La variance de l'estimateur $\hat{f}$ mesure à quel point les estimations peuvent varier autour de leur espérance. Elle est définie par :

$$\mathrm{Var}_P(\hat{f}) = \mathrm{Var}_{O \sim P}[\hat{f}(O)]$$

5.3 Risque

Le risque est une mesure globale de la qualité de l'estimateur, prenant en compte à la fois le biais et la variance. Il est défini par :

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

où L est une fonction de coût.

On cherche à minimiser ce risque.

6 Estimateur Admissible

Un estimateur est dit admissible s'il n'existe pas d'autre estimateur qui le surpasse en termes de risque. Dans la recherche d'un estimateur admissible, il est nécessaire de trouver un compromis entre le biais et la variance.

Un biais faible signifie que, en moyenne, nos estimations sont proches de la véritable valeur. Cependant, un estimateur non biaisé n'est pas nécessairement meilleur, il vaut mieux avoir un peu de biais que pas du tout. Inversement, un estimateur avec une faible variance n'est pas non plus idéal.