



MAT 3601a - Statistique

Généralités, précisions pour la suite

Relire la fiche de proba (SIC 3101) : les mêmes notations seront utilisées, et pas toujours redéfinies. On notera quand même $m_X = \mathbb{E}[X]$, $\sigma_X^2 = \mathbb{V}(X)$...

Revoir aussi l'algèbre bilinéaire (notamment les problèmes de minimisation de distances dans les espaces préhilbertiens).

Certaines définitions feront appel à la suite de la fiche. Du coup, lisez là plutôt deux fois qu'une..

Rappels et compléments de probabilités

Espérance (appelée aussi moyenne) : la base

Définition.

$$\text{Cas discret : } \mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{k \in \mathbb{Z}} kP(X = k)$$

$$\text{Cas continu : } \mathbb{E}(X) = \int_{\Omega} X(\omega)dP(\omega) = \int_{\mathbb{R}} x dP_X(x) = \left(\int_{\mathbb{R}} x f_X(x) dx \text{ si la densité } f_X \text{ existe} \right)$$

Remarques :

- L'espérance n'a donc de sens que si la v.a. X est intégrable sur \mathbb{R} , penser à le dire.
- Comme le montre la loi des grands nombres, l'espérance peut être interprétée comme la **limite de l'estimateur empirique** des v.a. (X_1, \dots, X_n) (défini par : $\hat{p}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$)
- $(\mathbb{E}[X - \mathbb{E}[X]])^2 = \inf_{\alpha} (\mathbb{E}[X - \alpha])^2$
- Pour les calculs, noter que si x est une constante/fonction déterministe/ n'est **pas** une v.a., alors $E[x] = x$. Noter que ce type d'objet est alors **indépendant** de toute v.a.

De plus, l'espérance est linéaire et l'espérance d'une indicatrice est une probabilité :

$$\mathbb{E}(\mathbb{1}_A) = P(\mathbb{1}_A = 1) = P(A)$$



Théorème de transfert

Proposition (Théorème du transfert).

$$\mathbb{E}(g(X)) \triangleq \int_{\mathbb{R}} g(x) dP_X(x) \left(= \int_{\mathbb{R}} g(x) f_X(x) dx \text{ si la densité } f_X \text{ existe} \right)$$

Sert très souvent, notamment pour calculer $\mathbb{E}[Y]$ sans la loi de Y .

En particulier : $\mathbb{V}(X) \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 dP_X(x)$

Erreur quadratique moyenne

Soit θ un paramètre que l'on cherche à **estimer**¹ grâce à une variable aléatoire $\hat{\theta}$ (**estimateur**).

On définit l'**erreur quadratique moyenne par** : $\text{eqm}(\hat{\theta}) \triangleq \mathbb{E}[(\hat{\theta} - \theta)^2]$

Théorème : $\text{eqm}(\hat{\theta}) = \mathbb{V}(\hat{\theta}) + b^2(\hat{\theta})$, où $b(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$ est le biais de $\hat{\theta}$.

La plupart du temps, on demande dans les exos de montrer que $\hat{\theta}$ est sans biais (i.e que $b(\hat{\theta}) = 0$), puis qu'on a convergence en moyenne quadratique (de $\hat{\theta}_n$ vers θ) : il suffit alors, avec cette formule, de montrer que $\lim_{n \rightarrow \infty} \mathbb{V}(\hat{\theta}_n) = 0$.

Variance et covariance

- $\text{Cov}(X, Y) \triangleq \mathbb{E}[(X - m_X)(Y - m_Y)] = \mathbb{E}[XY] - m_X m_Y$ pour les calculs.
- $\mathbb{V}(X) = \text{Cov}(X, X) \triangleq \mathbb{E}[(X - m_X)^2] = \mathbb{E}[X^2] - m_X^2$ pour les calculs.

Conditionnement : $\mathbb{E}[Y/X = x] = \int_{\mathbb{R}} y (f_Y^{X=x}(y)) dy = \int_{\mathbb{R}} y \left(\frac{f_{(X,Y)}(x, y)}{f_X(x)} \right) dy$

- $\mathbb{E}[\phi(X, Y)] = \mathbb{E}[\mathbb{E}[\phi(X, Y) / \underset{\text{connu}}{Y}]]$ d'où le cas particulier :
- $\mathbb{E}[U] = \mathbb{E}[\mathbb{E}[U / \underset{\text{connue}}{V}]] = \mathbb{E}[U/V = F]P(F) + \dots + \mathbb{E}[U/V = G]P(G)$

Remarque : $\mathbb{E}[(U - \mathbb{E}[U/V])^2] = \inf_{\phi} \mathbb{E}[(U_{\phi}(V))^2]$

1. C'est un peu l'objet du chapitre... Ce sera toujours le but du "**Jeu**".



Différents types de convergence

Loi des grands nombres

Proposition (Loi **forte** des grands nombres). Soit (X_i) une suite de v.a. décorrelées (souvent carrément indépendantes) et de moyenne commune $\mathbb{E}[X] = \mathbb{E}[X_1] = \dots = \mathbb{E}[X_n]$:

$$(\mathbb{E}(X) < \infty) \Rightarrow \left(\hat{p}(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{P.S.} \mathbb{E}(X) \right)$$

Théorème centrale limite

Soit (X_i) une suite de v.a.i.i.d telle que : $\begin{cases} \forall i \in \mathbb{N}, \mathbb{E}(X_i) = m \\ \forall i \in \mathbb{N}, V(X_i) = \sigma^2 \end{cases}$; alors on a :

$$\frac{\hat{p}(X_1, \dots, X_n) - m}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{Loi} \mathcal{N}(0, 1), \text{ i.e. : } \hat{p}(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{Loi} \mathcal{N}(m, \sigma^2/n)$$

En pratique, on approxime la loi de \hat{p} par cette gaussienne dès que $n \geq 30$.

Convergence en loi² (ou convergence étroite) : $X_n \xrightarrow[n \rightarrow \infty]{Loi} X \Leftrightarrow \begin{cases} F_{X_n} \xrightarrow[n \rightarrow \infty]{} F_X \\ f_{X_n} \rightarrow f_X \end{cases}$

(Il suffit que les points à droite de l'équivalence soient valable en tout point où l'on a continuité.)

Convergence en probabilité : $X_n \xrightarrow[n \rightarrow \infty]{P.} X \Leftrightarrow \begin{cases} \forall \varepsilon > 0, P(\|X_n - X\| \leq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 1 \\ \text{i.e} \\ \forall \varepsilon > 0, P(\|X_n - X\| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0 \end{cases}$

Convergence presque sûrement³ : $X_n \xrightarrow[n \rightarrow \infty]{P.S.} X \Leftrightarrow P(\lim_{n \rightarrow \infty} X_n = X) = 1$

Remarque. $Cv \text{ p.s.} \Rightarrow Cv \text{ en proba} \Rightarrow Cv \text{ en loi}$.

Proposition (Inégalité de Bienaymé-Tchebichev).

$$\forall \lambda \in \mathbb{R}, P(|X - m_X| \geq \varepsilon) \leq \left(\frac{\sigma_X}{\varepsilon} \right)^2, \text{ i.e. : } \forall \lambda \in \mathbb{R}, P(|X - m_X| \geq \varepsilon \sigma_X) \leq \frac{1}{\varepsilon^2}$$

Dans la suite de la pougne commence le vrai cours de Statistiques. On va utiliser tous les outils précédemment introduits pour estimer un paramètre θ par des estimateurs $\hat{\theta}$.

2. En gros, cela signifie que la convergence s'applique aux lois de probabilités.
3. Équivaut à une convergence simple sauf sur un ensemble de mesure nulle.



Estimation paramétrique

On introduit une suite $(\hat{\theta}_n)_{n \in \mathbb{N}^*}$ pour approcher θ .

Méthodes de construction des estimateurs paramétriques

Estimateurs sans biais :

Ici, on s'arrange pour qu'après calcul, $\boxed{b(\hat{\theta}_n) = 0 \text{ i.e } \mathbb{E}[\hat{\theta}_n] = \theta}$.

Attention, cette propriété est insuffisante pour garantir la qualité d'un estimateur : un estimateur non biaisé peut être de moindre qualité par rapport à un estimateur biaisé⁴.

Méthode des moments⁵

En pratique, dans le cas où $\mathbb{E}[X] = f(\theta)$ et que l'on peut estimer $\mathbb{E}[X]$ par \hat{p} :

— On exprime θ en fonction de la moyenne : $\theta = f^{-1}(\mathbb{E}[X])$

— Dans l'expression précédente, on remplace la moyenne par son estimation : $\boxed{\hat{\theta}_n = f^{-1}(\hat{p})}$

On généralise en remplaçant "moyenne" par "moment" dans ce qui précède.

Méthode du maximum de vraisemblance

On observe n v.a $(X_i)_{i=1}^n$ **indépendantes**, de même loi f_θ .

Pour approcher θ , on va fixer des réalisations $(x_i)_{i=1}^n$ et définir $\hat{\theta}^{MV}$ comme le maximum de la loi jointe $f_\theta^n(x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$ (appelée vraisemblance, vue comme une fonction de θ) :

$$\boxed{\hat{\theta}^{MV} \text{ est donc solution des équations } \frac{df_\theta(x_1, \dots, x_n)}{d\theta} = 0 \Leftrightarrow \frac{d \text{Log} f_\theta(x_1, \dots, x_n)}{d\theta} = 0}$$

(par croissance du logarithme, si besoin de transformer le produit en somme.)

Méthode des moindres carrés

D'après le prof, cette méthode ne tombe traditionnellement pas en exam.

Regardez dans le poly si vous avez le temps.

4. Cela est d'autant plus clair que cela est relatif au critère de qualité choisi pour évaluer l'estimateur. Il n'y a pas que l'erreur quadratique moyenne...

5. Voir p.35 pour l'explication du poly et les définitions des moments. Noter que moyenne = moment d'ordre 1.



POUGNES PHOENIX TSP

Estimation Bayesienne : la semaine prochaine