ENS de Lyon $\qquad$ 2nd semestre 2020-2021

L3 - Intégration et Probabilités

TD 8

Convergence en variation totale, vecteurs gaussiens

Convergence en variation totale

On définit la distance en variation totale de deux mesures de probabilité sur $\mathbb{R}^d$ ainsi :

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\mu(A) - \nu(A)|$$

Exercice 1 *Variation totale et convergence en loi de variables discrètes.*

1. Justifier que

$$\|\mu - \nu\|_{TV} = 2 \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\mu(A) - \nu(A)| = \sup_{\substack{f:\mathbb{R}^d \to [-1,1] \\ f \text{ mesurable}}} |\mu(f) - \nu(f)| = \sup_{\substack{f:\mathbb{R}^d \to [-1,1] \\ f \text{ continue}}} |\mu(f) - \nu(f)|$$

2. Montrer que la convergence en variation totale implique la convergence en loi. Donner un exemple de convergence en loi qui n'est pas en variation totale.
3. Montrer que pour des variables aléatoires dans $\mathbb{Z}$, la convergence en loi est équivalente à la convergence en variation totale.
4. Soit $(X_n)_n$ une suite de variables aléatoires dans $\mathbb{Z}$, telle que pour tout $k \in \mathbb{Z}$, $\mathbb{P}(X = k)$ converge. Est-ce que X_n converge en loi ?

Exercice 2 *Variation totale et couplages.*

On reprend les notations de l'exercice précédent.

1. On suppose que μ, ν sont à densités par rapport à une mesure m sur $\mathbb{R}^d$, de densités respectives f, g. Montrer que

$$\|\mu - \nu\|_{TV} = \int_{\mathbb{R}^d} |f(x) - g(x)| dm(x) = 2 \left[1 - \int_{\mathbb{R}^d} \min(f(x), g(x)) dm(x)\right].$$

   Remarquons que l'hypothèse de cette question n'est pas restrictive, puisque μ et ν sont toujours à densité par rapport à $\mu + \nu$.

2. Soit $Z = (X, Y)$ une v.a. dans $\mathbb{R}^d \times \mathbb{R}^d$, telle que $X \sim \mu$ et $Y \sim \nu$. Montrer que

$$\|\mu - \nu\|_{TV} \leq 2\mathbb{P}(X \neq Y)$$

3. Montrer que l'on peut construire une variable aléatoire $Z = (X, Y)$ dans $\mathbb{R}^d \times \mathbb{R}^d$ de lois marginales μ, ν telle que

$$2\mathbb{P}(X \neq Y) \leq \|\mu - \nu\|_{TV}$$

4. En déduire que

$$\|\mu - \nu\|_{TV} = 2 \inf\{\mathbb{P}(X \neq Y) \mid (X, Y) \text{ v.a. dans } \mathbb{R}^d \times \mathbb{R}^d, X \sim \mu, Y \sim \nu\}$$

Exercice 3 *Inégalité de Le Cam.*

1. Soit $p \in [0, 1]$. Montrer qu'il existe un couple de variables aléatoires (X, Y) tel que X suit une loi de Bernoulli de paramètre p et Y suit une loi de Poisson de paramètre p et $\mathbb{P}(X \neq Y) \leq p^2$
2. Soit $(p_1, ..., p_n)$ une suite de réels de $[0, 1]$. Soit S_n la somme de n variables aléatoires indépendantes de lois de Bernoulli de paramètres $p_1, ..., p_n$. On note μ_n la loi de S_n. On note μ la loi de Poisson de paramètre λ où

$$\lambda = \sum_{i=1}^{n} p_i;$$

   Montrer que

$$\|\mu_n - \mu\|_{TV} \leq \sum_{i=1}^{n} p_i^2$$

3. En déduire une majoration de la vitesse de convergence (pour la distance de variation totale) de $\mathrm{Bin}(n, \lambda/n)$ vers $\mathcal{P}(\lambda)$.

Vecteurs gaussiens

Exercice 4 *Vecteurs gaussiens*

1. Soit Y un vecteur gaussien de dimension n, centré et de matrice de covariance $\sigma^2 I_n$, et A une matrice orthogonale. Quelle est la loi de AY ?
2. Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$ et T indépendante de X de loi $\frac{1}{2}(\delta_1+\delta_{-1})$. Montrer que $Y=TX$ suit une loi $\mathcal{N}(0,1)$. Le vecteur (X,Y) est-il gaussien ?
3. Soit (X,Y) un vecteur gaussien centré de matrice de covariance $\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}$ avec $\rho\in]-1,1[$. Quelle est la loi de $(\frac{1}{\sqrt{2}}(X+Y),\frac{1}{\sqrt{2}}(X-Y))$?
4. Soit $(X,Y,Z)\in\mathbb{R}^3$ un vecteur gaussien. On pose $U=X+Y+Z$ et $V=X-Y$.
   (a) Montrer que $(U,V)\in\mathbb{R}^2$ est gaussien.
   (b) A quelle condition sur la matrice de covariance de (X,Y,Z) les variables U et V sont-elles indépendantes ?

Exercice 5 *Théorème de Cochran et statistique des échantillons gaussiens*

Soit $X\sim\mathcal{N}_d(0,\sigma^2 I_d)$ vecteur gaussien. Soit $\mathbb{R}^d=E_1\bigoplus E_2$ une décomposition orthogonale de R^d avec $n_i=\dim(E_i)$. Soit $X_i=p_i(X)$ où p_i est la projection orthogonale sur E_i.

1. Justifier qu'il existe U matrice orthogonale telle que $p_i=UI^i_{n_i}U^*$ avec

$$I^1_{n_1}=\begin{pmatrix}1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0\end{pmatrix}\qquad I^2_{n_2}=\begin{pmatrix}0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1\end{pmatrix}$$

2. Quelle est la loi de U^*X ?
3. Montrer que X_1 et X_2 sont indépendantes et déterminer la loi de $\|X_i\|^2$.
4. Soit $(X_1,...,X_n)$ un échantillon de variables aléatoires iid de loi $\mathcal{N}(\mu,\sigma^2)$. On note

$$\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i\qquad S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$$

   Montrer que $\bar{X}$ et S^2 sont indépendantes et de lois respectives $\mathcal{N}(\mu,\frac{\sigma^2}{n})$ et $\sigma^2\frac{\chi^2(n-1)}{n-1}$. *En statistiques, $\bar{X}$ est une "estimation empirique" de la moyenne et S^2 une "estimation empirique" de la variance.*

5. On note $T(n)$ la loi de $\frac{N}{\sqrt{Z/n}}$ où N et Z sont indépendantes de lois respectives $\mathcal{N}(0,1)$ et $\chi^2(n)$. Construire une suite de variables aléatoires T_n à partir de $\bar{X},\mu,S$ de loi $T(n-1)$. *En statistiques, la connaissance des lois de $\frac{n-1}{\sigma^2}S^2$ et de T_n permet de construire des intervalles de confiance pour les "estimations empiriques"*