# Tricks espérance/conditionnelle

## 0.1 Rappels premier semestre:

- Théorème Central Limite : Soit $(X_n)_{n\geq 1}$ v.a. réelles indépendantes de même loi dans $L^2$ avec variance $\sigma^2$. Alors:
$$\frac{1}{\sqrt{n}}(X_1 + \cdots + X_n) - n \cdot \mathbb{E}[X_1]) \to^d_{n\to\infty} \mathcal{N}(0, \sigma^2)$$
- Loi forte des grands nombres : Soient $(X_n)_{n\geq 1}$ des v.a. indépendantes, de même loi, dans $L^1$. Alors:
$$\frac{1}{n}(X_1 + \cdots + X_n) \to_{n\to\infty} \mathbb{E}[X_1]$$

## 0.2 Espérance

- (Statistician dream) Soit X v.a. à densité $f(x)$ , alors :
$$\mathbb{E}(g(X)) = \int_\Omega g(x)f(x)dx$$
- (Cauchy-Schwarz) $\mathbb{E}(|XY|) \leq \mathbb{E}(X^2)^{\frac{1}{2}} \cdot \mathbb{E}(Y^2)^{\frac{1}{2}}$
- (Biessaymé-Tchebyshev-Markov) Soit $f : \mathbb{R} \to \mathbb{R}_+$croissante et X une v.a. réelle, alors $\forall a \in \mathbb{R}$ on a :
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(f(X))}{f(a)}$$
- (Markov) $X \geq 0\,, a > 0 \Rightarrow \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$
- (Tchebyshev) $X \in L^2, a > 0 \Rightarrow \mathbb{P}(|X - \mathbb{E}(X)| > a) \leq \frac{Var(X)}{a^2}$
- (Jensen) Soit $f : \mathbb{R} \to \mathbb{R}$ convexe positive, $X \in L^1$ alors :
$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$$
- (Chernov) $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(exp(\lambda X)}{exp(\lambda a)}$ $\forall \lambda > 0$
- $\mathbb{E}(g(X)) = g(0) + \int_0^\infty g'(y) \cdot \mathbb{P}(X \geq y)dy$
- Si X à valeurs dans $\mathbb{N}$, $\mathbb{E}(X) = \sum_{k=1}^\infty \mathbb{P}(X \geq k)$

## 0.3 Espérance conditionnelle (+check s.2 Martingales)

- Loi des probas totales : $\Omega = \bigcup\limits_{i=1}^{n} B_i$, $B_i \cap B_j = \varnothing$ , on a :
$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$
- Théorème de Bayes : si $\mathbb{P}(A), \mathbb{P}(B) > 0$ , alors on a :
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$
- Exemple dans le cas discret: $Y$ une VA à valeurs discrètes et $\mathcal{B} = \sigma(Y) \implies \mathcal{B}$ est atomique et on a:
$$\mathbb{E}[X \mid Y] = \sum_{y\in E'} \mathbb{1}_{Y=y}\mathbb{E}[X \mid Y = y]$$
Où $E' = \{y \in E \mid \mathbb{P}(Y = y) > 0\}$

- Definition si $B$ est un événement ($B \in \mathcal{A}$):

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}[X \cdot \mathbb{1}_B]}{\mathbb{P}(B)}$$

- Soit $\mathcal{B}$ sous tribu de $\mathcal{A}$, $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, il existe une unique variable aléatoire $\mathbb{E}(X|\mathcal{B}) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ telle que pour toute variable aléatoire $Z$ $\mathcal{B} - mesurable$ bornée on a :

$$\mathbb{E}(X \cdot Z) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})Z)$$

- Si $X$ est $\mathcal{B} - mesurable$, $\mathbb{E}(X|\mathcal{B}) = X$
- L'application $X \to \mathbb{E}(X|\mathcal{B})$ est linéaire.
- $\mathbb{E}(\mathbb{E}(X|\mathcal{B})) = \mathbb{E}(X)$
- $X \geq X' \Rightarrow \mathbb{E}(X|\mathcal{B}) \geq \mathbb{E}(X'|\mathcal{B})$ p.s.
- (Jensen) Soit $f : \mathbb{R} \to \mathbb{R}$ convexe positive, $X \in L^1$ alors :

$$\mathbb{E}(f(X)|\mathcal{B}) \geq f(\mathbb{E}(X|\mathcal{B}))$$

- Soit X v.a. réelle et Y v.a. $\mathcal{B} - mesurable$, alors si X et Y sont positives *ou* si X et $XY \in L^1$, on a :

$$\mathbb{E}(YX|\mathcal{B}) = Y\mathbb{E}(X|\mathcal{B})$$

- (Concaténation) Si $\mathcal{B}_1, \mathcal{B}_2$ sous tribus de $\mathcal{A}$ t.q. $\mathcal{B}_1 \subset \mathcal{B}_2$, alors $\forall X$ v.a. positive *ou* intégrable, on a :

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{B}_1)\,(= \mathbb{E}(\mathbb{E}(X|\mathcal{B}_1)|\mathcal{B}_2))$$

- (Indépendance) Deux sous-tribus $\mathcal{B}_1$et $\mathcal{B}_2$ sont indépendantes ssi $\forall X$ v.a. $\mathcal{B}_2 - mesurable$ positive (ou dans $L^1(\Omega, \mathcal{B}_2, \mathbb{P})$) on a :

$$\mathbb{E}(X|\mathcal{B}_1) = \mathbb{E}(X)$$

- Soit $X$ v.a. $\mathcal{B}_2 - mesurable$ positive, $\mathcal{B}_1$et $\mathcal{B}_2$ sous-tribus indépendantes, alors $\forall Z$ $\mathcal{B}_1 - mesurable$ positive on a :

$$\mathbb{E}(ZX) = \mathbb{E}(Z)\mathbb{E}(X) = \mathbb{E}(Z\mathbb{E}(X))$$

- Soit $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{A}$, on a pour $X, Y \in L^2$:
  - $\mathbb{E}(X|\mathcal{B}_2)^2 = \mathbb{E}(X\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_2) = \mathbb{E}(\mathbb{E}(X\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_2))$
  - $\mathbb{E}((X - \mathbb{E}(X|\mathcal{B}_2))^2 + \mathbb{E}((\mathbb{E}(X|\mathcal{B}_2) - \mathbb{E}(X|\mathcal{B}_1))^2 = \mathbb{E}((X - \mathbb{E}(X|\mathcal{B}_1))^2$
  - $Var(X) = \mathbb{E}[Var(X|\mathcal{B}_1)] + Var(\mathbb{E}[X|\mathcal{B}_1])$ où $Var(X|\mathcal{B}_1) = \mathbb{E}[X^2|\mathcal{B}_1] - \mathbb{E}[X|\mathcal{B}_1]^2$
- $\mathbb{P}(X_2 = x_2, ..., X_n = x_n) = \sum_{x_1 \in E} \mathbb{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$
- $\mathbb{P}(X_{k(n+1)} = y|X_0 = x_0, .., X_{kn} = x_n) = \mathbb{P}(X_{k(n+1)} = y|X_0 = x_0, .., X_{kn} = x_n, X_{k(n+1)-1} = x_{n+1})\mathbb{P}(X_{k(n+1)-1} = x_{n+1}|X_0 = x_0, , ..., X_{kn} = x_n)$

Calcul d'espérence conditionnel dans un cas fini:

$$\mathbb{E}[X \mid \mathcal{B}] = \sum_{A \text{ atomes de } \Omega} \mathbb{1}_A \cdot b_A, \qquad b_A := \begin{cases} \frac{\mathbb{E}[\mathbb{1}_A \cdot X]}{\mu(A)} & si \mu(A) \neq 0 \\ 0 & sinon \end{cases}$$

## 0.4 Temps d'arrêts:

$$\{T = n\} \in \mathcal{F}_n \iff \{T \leq n\} \in \mathcal{F}_n$$
$$\{T \wedge S \leq n\} = \{T \leq n\} \cup \{S \leq n\}$$
$$\{T \vee S \leq n\} = \{T \leq n\} \cap \{S \leq n\}$$
$$\{T + S = n\} = \bigcup_{k=0}^{n} \{T = k\} \cap \{S = n - k\}$$

## 0.5 Chaînes de Markov

**Quelques notations:**

$$(Qf)(x) = \sum_{y \in E} Q(x,y)f(y)$$

$$(\mu Q)(x) = \sum_{y \in E} \mu(y)Q(y,x)$$

Une chaine de Markov est:

- Irreductible si $\forall x, y \in E: \quad U(x,y) > 0 \implies \exists n \in \mathbb{N} \ : \ Q^n(x,y) > 0$
- Une CM est <u>récurrente positive</u> si existe mesure de proba invariante $\mu$:

$$\mathbb{E}_x[H_x] = \frac{1}{\mu(x)}$$

- Une CM est <u>récurrente nulle</u> si toute mesure de proba invariante $\mu$ à masse infinie:

$$\mathbb{E}_x[H_x] = \infty$$

- Une CM irréductible récurrente positive et apériodique (période = 1) est dite ergodique.

**Propriétés:**

- Soit $X_n$chaîne de Markov alors $|X_n|$ et $max\{X_k : 0 \leq k \leq n\} - |X_n|$ sont des C.M. mais pas $max\{X_k : 0 \leq k \leq n\}$
- $Q^{n+1}(x,y) = \sum_{z \in E} Q^n(x,z)Q(z,y)$ !!! $Q^n$ représente les chemins de taille n !!!
- $X_n$ est une chaîne de Markov $\iff$ $\mathbb{P}(X_0 = x_0, ..., X_n = x_n) = \mathbb{P}(X_0 = x_0)Q(x_o, x_1) \cdots Q(x_{n-1}, x_n)$
- Si $f : E \to \mathbb{R}+$ est mesurable, alors :

$$\mathbb{E}(f(X_{n+1})|X_0, .., X_n) = \mathbb{E}(f(X_{n+1})|X_n) = Qf(X_n)$$

- $\mathbb{P}(X_{n+1} = y_1, .., X_{n+p} = y_p | X_0 = x_0, .., X_n = x_n) = Q(x_n, y_1)Q(y_1, y_2) \cdots Q(y_{p-1}, y_p)$
- $\mathbb{P}_x(X_n = y) = Q^n(x,y)$ et $\mathbb{P}_\mu = \sum_{x \in E} \mu(x)\mathbb{P}_x$
- (Markov forte) Soit T temps d'arrêt fini p.s. , $F, G : \Omega \to \mathbb{R}$ borné avec $G - F_t \, mesurable$ , alors :

$$\mathbb{E}_\mu(1_{T<\infty}F(G \circ \theta_T)) = \mathbb{E}_\mu(1_{T<\infty}F\mathbb{E}_{X_T}(G))$$

- $U(x,y) = \mathbb{E}_x(N_y) \underset{\text{prop}}{=} \sum_{n=0}^{\infty} Q^n(x,y)$
- Si $x \neq y$ , $U(x,y) = \mathbb{P}_x(H_y < \infty)U(y,y)$
- $x$ est récurrent $\iff U(x,x) = \infty$
- Soit $x \in R$ et $y \in E$ tel que $U(x,y) > 0$ , alors $y \in R$ et $\mathbb{P}_y(H_x < \infty)$ et $U(y,x) > 0$
- (Classification des états) Il existe une partition de R, $R = \bigcup_{i \in I} R_i$ telle qu'on a :
  - Si $x \in R$, $i \in I$ est tel que $x \in R_i$, on a $\mathbb{P}_x$ p.s. :
    * $N_y = \infty \ \forall y \in R_i$
    * $N_y = 0 \ \forall y \in E \backslash R_i$
  - Si $x \in E \backslash R$, $T = inf\{n \geq 0 : X_n \in R\}$, on a $\mathbb{P}_x$ p.s. :
    * ou bien $T = \infty$ et $N_y < \infty \ \forall y \in E$

    * ou bien $T < \infty$ et il existe un indice (aléat) $j \in I$ tel que : $\forall n \geq T$, $X_n \in R_j$

- Si la chaîne est irréductible (ie $U(x,y) > 0$) alors :
  - Soit tous les états sont récurrents, il existe une seule classe de récurrence et on a pour tout $x \in E$ :
  $$\mathbb{P}_x(N_y = \infty, \forall y \in E) = 1$$
  - Soit tous les états sont transitoires et alors, $\forall x \in E$ :
  $$\mathbb{P}_x(N_y < \infty, \forall y \in E) = 1$$

  Et si E est fini, seul le cas 1 se produit.

- !!!récurrence marche aléatoire!!! Soit $Y_n = Y_0 + \sum_{i=1}^{n} \xi_i$ est une marche aléatoire, supposons que $\mathbb{E}[|\xi_1|] < \infty$ et soit $m = \mathbb{E}[\xi_1]$
  - Si $m \neq 0$, alors tous les états sont transitoires.
  - Si $m = 0$, tous les états sont récurrents. De plus, la chaîne est irréductible ssi le sous-groupe engendré par $\{y \in \mathbb{Z} : \mu(y) > 0\}$ est $\mathbb{Z}$ tout entier.
- La marche aléatoire simple sur un graphe fini connexe est récurrente irréductible.
- Réversible $\Rightarrow$ invariante ($\mu Q = \mu$)
- (construction mesure invariante) Soit $x \in R$ , la formule : $\mu(y) = \mathbb{E}_x\left[\sum_{k=0}^{H_x-1} 1_{\{X_k=y\}}\right]$ est une mesure invariante et si la chaîne est récurrente irréductible, elle est unique à constante multiplicative près. De plus, $\mu(y) > 0$ $\Leftrightarrow$ $y$ appartient à la classe de récurrence. (Aussi: $\mu(x) = 1$)
- ($\mathbb{E}_x(H_x)$ ) Supposons une chaîne récurrente irréductible. Alors :
  - Soit il existe une *mesure de probabilité* invariante $\mu$, et on a $\forall x \in E$: (récurrente positive)
  $$\mathbb{E}_x(H_x) = \frac{1}{\mu(x)}$$
  - Soit toute mesure invariante a une masse totale infinie et on a $\forall x \in E$: (récurrente nulle)
  $$\mathbb{E}_x(H_x) = \infty$$

  En gros on essaye de trouver une mesure réversible (donc invariante) avec la déf et ensuite on regarde les cas.
- Si on a une chaîne irréductible et une mesure invariante finie, alors la chaîne est récurrente (positive).

**Exemple important:**

La marche aléatoire sur un graph quelconque:
$$Q(x,y) = \frac{1}{\deg(x)} \mathbb{1}_{x \sim y}$$
$x \sim y$ en gros quand $x$ est en relation avec $y$, en quelque sorte, voisin. Une mesure invariante est alors:
$$\tilde{\mu}(x) := \deg(x) \implies \mu(x) := \frac{\tilde{\mu}(x)}{\sum_{z \in E} \tilde{\mu}(z)}$$
$\mu$ est une mesure de proba du coup

**Relations tirés d'exos:**
$$\mu \text{ reversible pour } Q \iff Q \text{ auto-adjoint sur } L^2(E, \mu)$$

## 0.6 Comportement asymptotique et période

Considérons la chaîne de Markov canonique associée à une matrice de transition Q.

- (comportement asymptotique) Supposons la chaîne récurrente irréductible, et soit $\mu$ une mesure invariante. Soient f et g fonctions positives sur E telles que $\int f d\mu < \infty$ et $0 < \int g d\mu < \infty$. Alors, pour tout $x \in E$, on a $\mathbb{P}_x$ p.s. :
$$\frac{\sum_{i=1}^{n} f(X_i)}{\sum_{i=1}^{n} g(X_i)} \underset{n\to\infty}{\longrightarrow} \frac{\int f\ d\mu}{\int g\ d\mu}$$
- Si la chaîne de Markov est irréductible et récurrente positive, et si $\mu$ désigne l'unique probabilité invariante, on a $\mathbb{P}_x$ p.s. :
$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to_{n\to\infty} \int f d\mu$$
- Supposons la chaîne récurrente irréductible, loi de $X_0$ arbitraire. Alors, pour tout $x \in E$, on a :
  1. dans le cas récurrent positif,
  $$\frac{1}{n}\sum_{k=0}^{n-1} 1_{X_k=x} \to^{p.s.}_{n\to\infty} \mu(x)$$
  où $\mu$ est l'unique probabilité invariante.
  2. dans le cas récurrent nul,
  $$\frac{1}{n}\sum_{k=0}^{n-1} 1_{X_k=x} \to^{p.s.}_{n\to\infty} 0$$
- $L_x := \{n \geq 0 : Q^n(x,x) > 0\}$ et $d(x) = PGDC(L_x)$ et $L_x$ est stable par addition.
Notons aussi que: $L_x - L_x = d(x)\mathbb{Z}$
- Supposons la chaîne récurrente irréductible,
  - Tous les points ont la même période, dite la période de la chaîne.
  - Si $d = 1$ (i.e. apériodique), alors pour tous $x, y \in E$, il existe un entier $n_0$ tel que $Q^n(x,y) > 0$ pour tout $n \geq n_0$.
- Supposons la chaîne irréductible, récurrente positive et apériodique. Alors, si $\mu$ désigne l'unique probabilité invariante, on a pour tout $x \in E$ :
$$\sum_{y\in E} |\mathbb{P}_x(X_n = y) - \mu(y)| \to_{n\to\infty} 0$$
(la loi de $X_n$converge vers $\mu$ en variation totale)

## 0.7 Statistique

- Une famille d'estimateurs $(T_n)_{n\in\mathbb{N}}$ de $f(\theta)$ est dite consistante si:
$$T_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta} f(\theta) \quad \forall\theta \in \Theta$$
- Le bias d'un estimateur est donné: $\mathbb{E}_\theta[T] - f(\theta) = \mathbb{E}_\theta[T - f(\theta)]$
- Si $\theta = \mathbb{E}_\theta[g(X_1)]$ alors un estimateur sans bias est:
$$\hat{\theta} := \frac{1}{n}(\ g(X_1) + \cdots + g(X_n)\ )$$
- Le risque quadratique est donné par la formule:
$$R_{\hat{\theta}}(\theta) := \mathbb{E}_\theta[\ (\hat{\theta} - \theta)^2\ ] = \mathrm{Var}_\theta(\hat{\theta}) + \Big(\mathbb{E}[\hat{\theta} - \theta]\Big)^2$$