Fiche de cours pour l'épreuve de modélisation (option A)

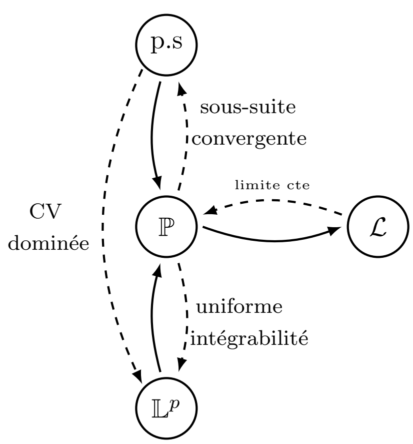
Sacha Quayle

1 Probabilités

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

Rappels de probabilités [App]

- Modes de convergence :
  1. $X_n \xrightarrow{ps} X$ si $\mathbb{P}(\{w \in \Omega,\, X_n(w) \to X(w)\}) = 1$.
  2. $X_n \xrightarrow{\mathbb{P}} X$ si : $\forall \epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$.
  3. $X_n \xrightarrow{\mathbb{L}^p} X$ si $\|X_n - X\|_p \to 0$.
  4. $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction f continue bornée, $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$.
- *Thm.* $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si l'une des conditions suivantes équivalentes est vérifiée :
  1. Pour tout t tel que F_X est continue en t, $F_{X_n}(t) \to F_X(t)$.
  2. *(Paul-Levy)* $(\varphi_{X_n})_n$ CVS vers φ_X.
- *Prop.* Si X_n à densité f_n et X à densité f tel que (f_n) CVS vers f, alors $X_n \xrightarrow{\mathcal{L}} X$ (réciproque fausse, cf contre-ex (1)).
- *Continuous mapping theorem.* Soit $X_n \xrightarrow{\mathcal{L}} X$ et f continue (p.s par rapport à $\mathbb{P}_X$.) Alors $f(X_n) \xrightarrow{\mathcal{L}} f(X)$.
- *Slutsky.* Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ constante, alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$, donc d'après le théorème précédent $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$ et $X_n Y_n \xrightarrow{\mathcal{L}} Xc$ (pas vrai si limite pas constante, cf contre-ex (2)).
- Articulation entre les différents modes de convergence (cf contre-ex (3), (4), (5)) :



- Soit $(X_n)_{n \in \mathbb{N}}$ VA i.i.d intégrables.
  1. *LfGN.* $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}(X_1)$.
  2. *LFGN.* $\frac{S_n}{n} \xrightarrow{\mathbb{L}^1,\, p.s} \mathbb{E}(X_1)$.
  3. *TCL.* Si $\mathbb{E}(|X_1|^2) < \infty$, en notant $\sigma = \sqrt{\mathbb{V}(X_1)}$. Alors $\sqrt{n}(\frac{S_n}{n} - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

Inégalités classiques

- *Inégalité de Bienaymé-Tchebychev.* Si X admet un moment d'ordre 2 fini, pour tout $x > 0$,
$$\mathbb{P}(|X - \mathbb{E}[X]| \geqslant x) \leqslant \frac{\mathrm{Var}[X]}{x^2}.$$
- *Inégalité de Hoeffding.* Si les variables X_i sont bornées $a_i \leqslant X_i \leqslant b_i$, alors pour tout $x > 0$,
$$\mathbb{P}(S_n - \mathbb{E}[S_n] \geqslant x) \leqslant \exp\left(-\frac{2x^2}{\sum (b_i - a_i)^2}\right).$$

Statistiques d'ordre [CR]

On considère n VA $(X_1, ..., X_n)$ indépendantes de même loi de densité f et de FDR F sur $\mathbb{R}$.

- *Prop.* Les (X_i) sont presque tous sûrement distincts. On pose $(X_{(i)})$ tels que $X_{(1)} < ... < X_{(n)}$. $X_{(i)}$ est apppelé statistique d'ordre de rang i de l'échantillon.
- *Prop.* Le n-uplet $(X_{(1)}, ..., X_{(n)})$ a pour densité
$$(x_1, ..., x_n) \mapsto n! 1_{x_1 < ... < x_n} \prod_{i=1}^{n} f(x_i).$$
- *Exemple.* Si les (X_i) suivent une loi uniforme sur $[0, t]$, alors $(X_{(1)}, ..., X_{(n)})$ a pour densité $(x_1, ..., x_n) \mapsto \frac{n!}{t^n} 1_{0 < x_1 < ... < x_n < t}$.

Vecteurs gaussiens [GK]

- Si $X_1, ..., X_n$ variables gaussiennes indépendantes, $(X_1, ..., X_n)$ est un vecteur gaussien (faux sans l'indépendance, cf contre-ex (6)).
- Si $(X_1, ..., X_n)$ vecteur gaussien, les VA sont indépendantes ssi elles sont non corrélées, i.e. la matrice de covariance est diagonale (faux dans le cas général, cf contre-ex (7)).
- Soit $X \sim \mathcal{N}_d(m, \Gamma)$. Alors $AX + b \sim \mathcal{N}_d(Am + b, A\Gamma {}^t A)$.
- *TCL multi-dimensionnel.* $(X_n)_n$ suite de VA à valeurs dans $\mathbb{R}^d$, i.i.d. et L^2. On note $\mu = \mathbb{E}[X_1]$, C sa matrice de covariance. Alors $\sqrt{n}(\frac{1}{n}\sum_{k=1}^n X_k - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}_d(0, C)$.
- *Théorème de Cochran.* Soit $X \sim \mathcal{N}_d(m, \sigma^2 I_d)$. Soit $\mathbb{R}^d = E_1 \oplus ... \oplus E_p$ décomposition en somme directe orthogonale avec $d_k = \dim E_k$. Soit Π_k la matrice de projection orthogonale sur E_k et $Y_k = \Pi_k X$. Alors :
  1. Les Y_k sont des vecteurs gaussiens indépendants et $Y_k \sim \mathcal{N}_d(\Pi_k m, \sigma^2 \Pi_k)$.
  2. Les VA $\|Y_k - \Pi_k m\|^2$ sont indépendantes et $\frac{\|Y_k - \Pi_k m\|^2}{\sigma^2} \sim \chi^2(d_k)$.

– *Corollaire.* Soit $(X_1, ..., X_n)$ un échantillon de loi $\mathcal{N}(m, \sigma^2)$. Alors $\overline{X_n}$ et $\hat{S}_n^2$ (cf définitions moyenne et variance empiriques) sont indépendantes, $\overline{X_n} \sim \mathcal{N}(m, \sigma^2/n)$ et $(n-1)\hat{S}_n^2/\sigma^2 \sim \chi^2(n-1)$.

– *Application.* Test du chi-deux, cf plus loin.

Simulation de variables aléatoires

– *Fonction de répartition inverse généralisé* d'une VA X de FDR F : $F^-(u) = \inf\{\, x \in \mathbb{R},\, F(x) \geqslant u \,\}$.

– Si $U \sim \mathcal{U}(0,1)$, $F^-(U)$ admet F pour fonction de répartition. On peut donc simuler la VA X en calculant F^-.

– *Méthode d'acceptation-rejet pour loi uniforme.* Soient $B \subset A \subset \mathbb{R}^d$ bornés, $(X_n)_n$ i.i.d. de loi $\mathcal{U}(A)$. Si $Y_n = \mathbf{1}_B(X_n)$ et $T = \inf\{\, n \geqslant 1,\, X_n \in B \,\}$, alors $X_T \sim \mathcal{U}(B)$.

– *Méthode d'acceptation-rejet général.* Soit f densité de proba et (X, Y) couple de loi uniforme sur $\{\, (x, y) \in \mathbb{R}^2,\, 0 < y < f(x) \,\}$ (simulé avec thm précédent). Alors $X \sim f$.

– *Méthode Box-Muller pour loi normale.* Soient R tel que $R^2 \sim \mathcal{E}(1/2)$ et $\theta \sim \mathcal{U}(0, 2\pi)$, VA indép. Alors $X = R\cos\theta$ et $Y = R\sin\theta$ sont indép. et de même loi $\mathcal{N}(0,1)$.

2 Martingales [CR]

Espérance conditionnelle

– *Déf espérance conditionnelle.* Si $\mathcal{B}$ sous-tribu de $\mathcal{F}$ et $X \in L^1$, $\mathbb{E}[X|\mathcal{B}]$ est l'unique variable Y intégrable et $\mathcal{B}$-mesurable telle que pour tout $B \in \mathcal{B}$, $\mathbb{E}[X 1_B] = \mathbb{E}[Y 1_B]$.

– Propriétés analogues à celles de l'espérance : linéarité, croissance, convergence monotone, lemme de Fatou, convergence dominée, Jensen.

– Propriétés spécifiques :
  1. L'espérance conditionnelle est intégrable et $\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X]$.
  2. Si X est $\mathcal{B}$-mesurable, $\mathbb{E}[X|\mathcal{B}] = X$.
  3. Si X, Z intégrables avec XZ intégrable et Z $\mathcal{B}$-mesurable, alors $\mathbb{E}[XZ|\mathcal{B}] = \mathbb{E}[X|\mathcal{B}]Z$.
  4. Si $\mathcal{B}_1 \subset \mathcal{B}_2$ sous-tribus, alors $\mathbb{E}[\mathbb{E}[X|\mathcal{B}_1]|\mathcal{B}_2] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_2]|\mathcal{B}_1] = \mathbb{E}[X|\mathcal{B}_1]$.
  5. Si $\sigma(X)$ indépendante de $\mathcal{B}$, alors $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$.

Martingales

– *Déf.* Une filtration est une suite croissante $(\mathcal{F}_n)$ de sous-tribus de $\mathcal{F}$. Un processus (X_n) est adapté (resp. prévisible) par rapport à cette filtration si X_n est $\mathcal{F}_n$-mesurable (resp. $\mathcal{F}_{n-1}$-mesurable). Si la filtration est implicite, on dira juste que le processus est adapté (resp. prévisible).

– *Exemple.* Si (X_n) processus, la filtration canonique définie par $\mathcal{F}_n = \sigma(X_0, ..., X_n)$ est une filtration.

– *Déf martingale.* (X_n) est une martingale (resp. sur-martingale, resp. sous-martingale) par rapport à $(\mathcal{F}_n)$ si, pour tout n, X_n est intégrable, (X_n) est adapté et $\mathbb{E}[X_{n+1}|\mathcal{F}_n] =$ (resp. $\leqslant$ resp. $\geqslant$) X_n.

– *Remarque.* Les martingales ont été introduites en lien avec les jeux de hasard, X_n représente la fortune du joueur à la n-ième partie et la propriété de martingale signifie que le jeu est équilibré.

– *Exemples.*
  1. Si X VA intégrable, alors $X_n = \mathbb{E}[X|\mathcal{F}_n]$ définit une martingale (martingale de Doob).
  2. Si (X_n) VA indépendantes centrées avec $X_0 = 0$, alors $S_n = X_0 + ... + X_n$ définit une martingale par rapport à la filtration canonique.
  3. Si (S_n) marche aléatoire de loi $p\delta_1 + (1-p)\delta_{-1}$, la marche est une martingale si $p = 1/2$, une sur-martingale si $p < 1/2$ et une sous-martingale si $p > 1/2$.

– *Prop (espérance constante).* Si (X_n) martingale, $\mathbb{E}[X_n] = X_0$.

– *Prop.* Si (X_n) martingale et φ convexe, $(\varphi(X_n))$ sous-martingale : application avec $\varphi(x) = |x|$ ou $\varphi(x) = x^2$.

Temps d'arrêt et convergence des martingales

– *Déf temps d'arrêt.* VA T à valeurs dans $\overline{\mathbb{N}}$ telle que pour tout $n \in \mathbb{N}$, $\{T = n\} \in \mathcal{F}_n$. Le processus arrêté associé à un processus adapté (X_n) est $X^{|T} = (X_{n\wedge T})$.

– *Exemples.* Temps constant $T = k$, temps d'atteinte.

– *Thm.* Si (X_n) martingale (resp. sur-martingale, resp. sous-martingale) et T temps d'arrêt, le processus arrêté est une martingale (resp. sur-martingale, resp. sous-martingale). Si T est presque sûrement fini, alors $X_{n\wedge T} \xrightarrow{ps} X_T$.

– *Théorème d'arrêt borné.* Si (X_n) martingale (resp. sur-martingale, resp. sous-martingale) et T temps d'arrêt borné par un entier N, alors $\mathbb{E}[X_0] = \mathbb{E}[X_T] = \mathbb{E}[X_N]$ (resp. $\mathbb{E}[X_0] \geqslant \mathbb{E}[X_T] \geqslant \mathbb{E}[X_N]$, resp. $\mathbb{E}[X_0] \leqslant \mathbb{E}[X_T] \leqslant \mathbb{E}[X_N]$).

– *Théorème d'arrêt non borné.* Si (X_n) martingale (resp. sur-martingale, resp. sous-martingale) et T temps d'arrêt presque sûrement fini tel que $\mathbb{E}[|X_T|] < \infty$ et $\mathbb{E}[1_{T>n}X_n] \to 0$, alors $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ (resp. $\mathbb{E}[X_0] \geqslant \mathbb{E}[X_T]$, resp. $\mathbb{E}[X_0] \leqslant \mathbb{E}[X_T]$).

– *Applications du thm d'arrêt.*

1. Si (X_n) sous-martingale positive, alors pour tout $\lambda > 0$, $\mathbb{P}\left(\max_{0\leqslant k\leqslant n} X_k > \lambda\right) \leqslant \dfrac{\mathbb{E}[X_n]}{\lambda}$ et $\mathbb{P}\left(\sup_{k\in\mathbb{N}} X_k > \lambda\right) \leqslant \dfrac{\sup \mathbb{E}[X_k]}{\lambda}$.
2. Si (X_n) sur-martingale positive, alors pour tout $\lambda > 0$, $\mathbb{P}\left(\max_{0\leqslant k\leqslant n} X_k > \lambda\right) \leqslant \dfrac{\mathbb{E}[X_0]}{\lambda}$.

– *CV p.s pour bornée L^1.* Si (X_n) martingale bornée dans L^1, alors (X_n) converge p.s (encore vrai pour sur ou sous-martingale).

– *CV p.s et L^2 pour bornée L^2.* Si (X_n) martingale bornée dans L^2, alors (X_n) converge dans L^2 et p.s. vers une VA X_∞ vérifiant $X_n = \mathbb{E}[X_\infty|\mathcal{F}_n]$.

– On peut démontrer la loi forte des grands nombres grâce aux martingales (cf [Ben])

3 Chaînes de Markov [CR] [Ben]

Définitions et propriétés

Soit E un ensemble fini ou dénombrable.

– *Déf chaîne de Markov.* (X_n) à valeurs dans E est une chaîne de Markov si pour tout $(x_0, ..., x_{k+1})$ dans E,

$$\mathbb{P}(X_{k+1} = x_{k+1}|X_0 = x_0, ..., X_k = x_k) \\ = \mathbb{P}(X_{k+1} = x_{k+1}|X_k = x_k).$$

On note $p_{x,y} = \mathbb{P}(X_{k+1} = y|X_k = x)$ et $P = (p(x,y))_{x,y\in E}$ la matrice de transition.

– *Exemples.* Chaîne à deux états, urne d'Ehrenfest, marche aléatoire sur $\mathbb{Z}$ ou $\mathbb{Z}^d$ (cf [CR]).

– *Relation de Chapman-Kolmogorov.* Si μ_k est la loi de X_k, alors $\mu_{n+1} = \mu_0 P^{n+1}$ et $\mathbb{P}(X_n = y|X_0 = x) = P^n(x,y)$. On note $\mathbb{P}_x(X_n = y) = \mathbb{P}(X_n = y|X_0 = x)$.

– *Propriété de Markov faible.* Conditionnellement à $X_n = x$, la suite $(X_{n+k})_k$ est une chaîne de Markov, indépendante de $X_0, ..., X_{n-1}$, de même matrice de transition et de loi initiale δ_x.

– *Propriété de Markov forte.* Si T temps d'arrêt, conditionnellement à $T < \infty$ et $T = i$, la suite $(X_{T+k})_k$ est une chaîne de Markov, indépendante de $X_0, ..., X_{n-1}$, de même matrice de transition et de loi initiale δ_{X_i}. On a donc, par exemple, la prop de Markov en remplaçant T par le temps d'atteinte (cf déf en dessous).

Classification des états

– *Déf.* Etats accessibles, états qui communiquent entre eux, classes, chaîne irréductible (cf [CR]).

– *Déf.* Période d'un état j : PGCD des entiers $n \geqslant 1$ tels que $p^n_{j,j} > 0$. Une chaîne apériodique est une chaîne dont tous les états sont apériodiques, i.e. de période 1.

– *Déf.* Temps d'atteinte de l'état j : $T_j = \inf\{n \geqslant 1,\ X_n = j\}$. Nombre de visites : $N_j = \sum_{n\geqslant 0} 1_{X_n=j}$.

– *Etat récurrent/transcient.* L'état j est récurrent si $\mathbb{P}_j(T_j < \infty) = 1$, transcient sinon. Il est récurrent si et seulement si la série $\sum_{n\geqslant 0} p^n_{j,j}$ converge.

– *Etat récurrent positif/nul.* Un état récurrent j est positif si $\mathbb{E}_j[T_j] < \infty$ et nul sinon.

– *Prop.* La périodicité, la transcience, la récurrence positive et la récurrence nulle sont des propriétés de classe. Par ex, si la chaîne est irréductible sur un espace d'états fini, elle est récurrente positive.

– *Exemple.* La marche aléatoire sur $\mathbb{Z}^d$ est récurrente pour $d = 1, 2$ et transitoire si $d \geqslant 3$.

Lois invariantes et théorème ergodique

– *Déf mesure/proba stationnaire.* Loi de mesure / probabilité μ telle que $\mu = \mu P$.

– *Théorème ergodique (cas fini).* Soit (X_n) chaîne de Markov sur espace d'états fini. Elle possède au moins une loi de proba invariante. Si de plus la chaîne est irréductible, il y a unicité et elle est donnée par

$$\mu(i) = \frac{1}{\mathbb{E}_i[T_i]}.$$

Si de plus la chaîne est est apériodique, alors quelle que soit la loi initiale, (X_n) converge en loi vers l'unique proba invariante.

– *Théorème ergodique (cas général).* Soit (X_n) chaîne irréductible récurrente. Alors la chaîne admet une mesure invariante, et toutes les mesures invariantes sont proportionnelles. Si de plus la chaîne est récurrente positive, il existe une unique loi de proba invariante, donnée par

$$\mu(i) = \frac{1}{\mathbb{E}_i[T_i]},$$

et pour toute fonction f μ-intégrable,

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \xrightarrow{ps} \int f \mathrm{d}\mu.$$

Si de plus la chaîne est apériodique, alors quelle que soit la loi initiale, (X_n) converge en loi vers l'unique proba invariante.

– *Remarques.*

1. Si P et ${}^tP$ sont stochastiques, alors la loi uniforme est invariante.
2. Si la chaîne possède une seule classe récurrente on a les mêmes conclusions. Si elle possède plusieurs classes récurrentes, sur chacune de ses classes il y a une unique loi de proba invariante, et toute combinaison convexe de ces probas est une mesure invariante.
3. Si la chaîne est périodique, il n'y a pas convergence en loi : ex si chaîne de période 2 et $X_0 \neq \mu$ suit la loi δ_1, $\mathbb{P}(X_n = 1) = 0$ pour n impair mais $\mathbb{P}(X_n = 1)$ ne tend pas vers 0.

4 Processus de Poisson [CR]

- *Déf processus de Poisson.* $(N(t))_t$ de paramètre $\lambda > 0$ tel que $N(0) = 0$ et
  1. Quels que soient $0 \leqslant t_1 \leqslant ... \leqslant t_k$, les VA $N(t_1), N(t_2) - N(t_1), ..., N(t_k) - N(t_{k-1})$ sont indépendantes (on dit alors que le processus est à accroissements indép.).
  2. Pour tout $s, t \geqslant 0$, $N(s+t) - N(t)$ suit une loi de poisson de paramètre λs.
- *Prop.* Un processus de Poisson est à accroissements stationnaires, i.e pour tout $s, t \geqslant 0$, la VA $N(t+s) - N(t)$ a même loi que $N(s)$.
- *Déf équivalente.* Processus de comptage associé aux instants (T_n) définis par $T_0 = 0$ et $T_{n+1} = T_n + \xi_n$ où (ξ_n) sont i.i.d de loi $\mathcal{E}(\lambda)$, i.e. $N(t) = \sum_{n \geqslant 1} 1_{T_n \leqslant t}$ (donne méthode pour simuler un processus de Poisson).
- *Prop.* On conserve les mêmes notations. Soit $(N(t))_t$ un processus de Poisson. Alors, conditionnellement à $\{N(t) = n\}$, le n-uplet $(T_1, ..., T_n)$ a même loi que la statistique d'ordre de n VA indépendantes de loi uniforme sur $[0, t]$.
- *Thm limites.* Soit $(N(t))_t$ un processus de Poisson de paramètre λ. Alors $\frac{N(t)}{t} \overset{ps}{\to} \lambda$ et $\sqrt{t}\left(\frac{N(t)}{t} - \lambda\right) \overset{\mathcal{L}}{\to} \mathcal{N}(0, \lambda)$.

5 Estimation [CR] [RS]

Modèles statistiques : premières définitions

On fixe $(\mathcal{X}, \mathcal{B}, \{P_\theta\}_{\theta \in \Theta})$ un modèle statistique. On cherche à estimer une fonction de θ, notée $\psi(\theta)$, sur la base d'un échantillon (i.e. une suite d'observations indépendantes et de même loi). Souvent, $\psi = Id$. On fixe n la taille de l'échantillon.

- Le modèle est dit
  1. *identifiable* si $\theta \mapsto P_\theta$ est injective.
  2. *paramétrique* si $\Theta \subset \mathbb{R}^d$ pour $d \in \mathbb{N}^*$.
  3. *d'échantillonnage* si $\mathcal{X} = \mathcal{X}^n$ et $P_\theta = p_\theta^{\otimes n}$ pour $n \in \mathbb{N}^*$.
- Une *statistique* est une fonction mesurable sur $(\mathcal{X}, \mathcal{B})$.
- Un *estimateur* de $\psi(\theta)$ sur $(\mathcal{X}, \mathcal{B})$ est une statistique à valeurs dans un sur-ensemble de $\psi(\Theta)$. On la note $\hat{\psi}$ (ou $\hat{\theta}$ si $\psi = Id$) et elle ne dépend pas de θ, seulement de l'échantillon $(X_1, ..., X_n)$.
- Si le modèle est paramétrique et d'échantillonnage, un *estimateur des moments* est une solution du système d'équations : $\forall i \in [\![1, d]\!]$, $\mathbb{E}_\theta[X_1^i] = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^j$.
- *Exemple.* Si le modèle statistique est $(\mathbb{N}^n, \mathcal{P}(\mathbb{N}^n), \{\mathcal{P}(\lambda)\}_{\lambda > 0})$, on a $\mathbb{E}[X_1] = \lambda$. Un estimateur par la méthode des moments de λ est donc $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$.
- La *mesure (ou loi) empirique* associée à un n-échantillon $(X_1, ..., X_n)$ est la mesure de probabilité $\hat{P}_n$ définie par $\hat{P}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n \delta_{X_i}$.
- La *moyenne empirique* est $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$.
- La *variance empirique* est $S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X_n})^2$.
- La *variance empirique corrigée* est $\hat{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2$.
- La *fonction de répartition empirique* est l'application $\hat{F}_n : \Omega \to \mathcal{F}(\mathbb{R}, [0,1])$ définie par :
  $\forall w \in \Omega$, $\forall t \in \mathbb{R}$, $\hat{F}_n(w)(t) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n \mathbf{1}_{X_i(w) \leqslant t}$.
- *Glivenko-Cantelli.* Soit $(X_n)_n$ de FDR F, et $(F_n)_n$ la suite des FDR empiriques associée. Alors : $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0$ p.s.
-
  1. Un modèle statistique est *dominé* par une mesure σ-finie ν si pour tout $\theta \in \Theta$, P_θ est absolument continue par rapport à ν.
  2. Une *vraisemblance* est une fonction mesurable $L : \mathcal{X} \times \Theta \to [0, +\infty[$ telle que
  $$\forall x \in \mathcal{X},\ \forall \theta \in \Theta,\ L(x, \theta) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(x).$$
  3. La *log-vraisemblance* est la fonction $l := \ln(L)$.
  4. Un *estimateur du maximum de vraisemblance* (EMV) de θ est une statistique T telle que
  $$\forall x \in \mathcal{X},\ T(x) \in \operatorname{argmax}_{\theta \in \Theta} L(x, \theta).$$
- Si la mesure dominante est la mesure de comptage, alors $L(x, \theta) = \displaystyle\prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i)$, si c'est la mesure de Lebesgue, alors $L(x, \theta) = \displaystyle\prod_{i=1}^n f_{X_i}(x_i)$.
- *Exemple.* Si le modèle statistique est $((\mathbb{R}_+)^n, \mathcal{B}((\mathbb{R}_+)^n), \{\mathcal{E}(\lambda)^{\otimes n}, \lambda > 0\})$, la vraisemblance associée est $L(x, \lambda) = e^{n \ln \lambda - \lambda \sum x_i}$, donc un EMV est $\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i}$.
- Un modèle statistique dominé est dit *exponentiel* de dimension $d' \in \mathbb{N}^*$ s'il existe deux fonctions $h : \Theta \to \mathbb{R}$ et $g : \Theta \to \mathbb{R}^{d'}$, et deux statistiques $T : (\mathcal{X}, \mathcal{B}) \to \mathbb{R}^{d'}$ et $c : (\mathcal{X}, \mathcal{B}) \to]0, +\infty[$ telles que
  $$L(x, \theta) = c(x) e^{\langle g(\theta), T(x) \rangle - h(\theta)}.$$
  On appelle T *statistique naturelle* et g *paramètre naturel.*

Estimateurs ponctuels : critères de performance

On fixe $(\mathcal{X}, \mathcal{B}, \{P_\theta\}_{\theta\in\Theta})$ un modèle statistique. On note m l'espérance commune, et σ^2 la variance.

- Un estimateur $\hat{\psi}$ est dit *sans biais* si : $\mathbb{E}_\theta[\hat{\psi}] = \psi(\theta)$. Il est dit asymptotiquement sans biais si $\mathbb{E}_\theta[\hat{\psi}] \to \psi(\theta)$.
- *Exemple.* La moyenne empirique est un estimateur sans biais de m. La variance empirique (resp. variance empirique corrigée) est un estimateur asymptotiquement biaisé (resp. sans biais) de σ^2.
- Un estimateur $\widehat{\psi} := T(X)$ est dit
  1. *faiblement consistant* si $\widehat{\psi} \xrightarrow{\mathbb{P}_\theta} \psi(\theta)$.
  2. *fortement consistant* si $\widehat{\psi} \xrightarrow{\mathbb{P}_\theta - ps} \psi(\theta)$.
- Soit $(v_n)_n$ une suite réelle positive qui tend vers $+\infty$. L'estimateur $\hat{\psi}$ est de vitesse v_n si pour tout $\theta \in \Theta$, il existe une loi λ_θ non dégénérée sur $\mathbb{R}^{d'}$ telle que
$$v_n(\hat{\psi} - \psi(\theta)) \xrightarrow{\mathbb{P}_\theta - loi} \lambda_\theta.$$
Si les λ_θ sont de lois normales, l'estimateur est dit *asymptotiquement normal.*
- *Méthode delta.* Si modèle paramétrique et la fonction ψ est de classe C^1., si $\hat{\theta}$ est un estimateur de θ de vitesse v_n, i.e
$$v_n(\hat{\theta} - \theta) \xrightarrow{\mathbb{P}_\theta - loi} \lambda_\theta.$$
alors
$$v_n(\psi(\hat{\theta}) - \psi(\theta)) \xrightarrow{\mathbb{P}_\theta - loi} J_{\psi(\theta)}\lambda_\theta,$$
où $J_{\psi(\theta)}$ est la matrice jacobienne de ψ au point θ.
- Si $d = 1$, et les X_i ont un moment d'ordre 2, d'espérance μ et de variance σ^2, alors
  1. La moyenne empirique est un estimateur fortement consistant et asymptotiquement normal de μ :
  $$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$
  2. La variance empirique est un estimateur fortement consistant de σ^2. Si les X_i ont de plus un moment d'ordre 4, alors il est asymptotiquement normal :
  $$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{Var}[(X_1 - \mu)^2])$$

Estimation par régions de confiance

On fixe $\alpha \in]0, 1[$ le niveau de risque.

- Une *région de confiance* de niveau de confiance $(1-\alpha)$ pour $\psi(\theta)$ est une famille $C = (C(x))_{x\in\mathcal{X}}$ de parties de $\psi(\Theta)$ telle que
$$\inf_{\theta\in\Theta} P_\theta(\psi(\theta) \in C(X)) \geqslant 1 - \alpha.$$
On dit alors qu'on a une région par excès. Lorsqu'on a égalité, on parle de niveau exactement égal à $1 - \alpha$. La construction se fait souvent directement à l'aide d'inégalités de probabilité.
- *Exemple.* Si $X_1, ..., X_n$ est un n-échantillon de loi $B(p)$, l'inégalité de Bienaymé-Tchebychev donne $\mathbb{P}(|\overline{X_n} - p| > \varepsilon) \leqslant 1/4n\varepsilon^2$ donc on déduit un intervalle de confiance.
- Une *région de confiance asymptotique* de niveau de confiance $(1 - \alpha)$ est une suite $(C_n)_n$ de régions de confiance pour $\psi(\theta)$ telle que
$$\underline{\lim}\, P_\theta(\{x \in \mathcal{X},\ \psi(\theta) \in C_n(x)\}) \geqslant 1 - \alpha.$$
La construction se fait à l'aide de théorèmes de convergence.
- *Exemple.* Si $X_1, ..., X_n$ est un n-échantillon de loi $\mathcal{P}(\lambda)$, on utilise le TCL, le fait que la moyenne empirique est aussi est un estimateur de la variance, donc de λ, et le lemme de Slutsky.
- Le *quantile d'ordre q* d'une loi de probabilité $\mathbb{P}$ est l'élément x_q tel que $\mathbb{P}(X \geqslant x_q) = 1 - q$.

6 Tests d'hypothèses [CR] [RS]

- *Principe des tests.*
  1. Définir l'hypothèse nulle H_0 et l'hypothèse alternative H_1.
  2. Choisir une statistique pour contrôler H_0.
  3. Définir la distribution de la statistique sous l'hypothèse H_0.
  4. Définir la région critique associée.
  5. Calculer la statistique.
  6. Prendre une décision.
- Il existe plusieurs types de tests :
  1. Test de conformité : confronter un paramètre calculé sur l'échantillon à une valeur pré-établie. Ex : tests sur la moyenne, la variance ou les proportions.
  2. Test d'adéquation (resp. adéquation à une famille de lois) : vérifie la compatibilité des données avec une distribution pré-établie (resp. famille de distributions).
  3. Test d'homogénéité : vérifie si deux échantillons proviennent de la même distribution.
  4. Test d'indépendance : vérifie l'indépendance entre deux échantillons.
- *Remarque.* Il y a une dissymétrie : on rejette H_0 ou on ne la rejette pas, dans aucun cas on l'accepte. Il faut voir un test statistique comme un raisonnement par l'absurde sur l'hypothèse H_0.
- *Zone de rejet.* Noté R_{H_0}, c'est l'ensemble des valeurs possibles pour l'observation conduisant au rejet de H_0 au profit de H_1. Souvent, R_{H_0} est de la forme $R_{H_0} = \{\, x,\ T(x) \leqslant s \,\}$ où T est la statistique de test et s est la valeur critique du test.
- *Risque de première espèce.* Risque de rejeter H_0 alors qu'elle est vraie. On la note α. Un test est dit de niveau (resp. taille) α_0 si : $\sup_{\theta\in\Theta} \alpha(\theta) \leqslant$ (resp. $=$) α_0.

- On se place dans un cas unilatéral, par exemple $(H_0) : \theta \leqslant \theta_0$ et $(H_1) : \theta > \theta_0$. Alors dans ce cas, la *p-valeur* est

$$p(x) = \sup_{\theta_0 \text{"}\in\text{"} H_0} P_{\theta_0}(T(X) \geqslant T(x)) = 1 - F_{T(X)}(T(x)).$$

Il s'agit de la valeur maximale du niveau pour laquelle on ne rejette pas H_0. Si $p(x) \leqslant \alpha$, on rejette H_0 au profit de H_1. Si $p(x) > \alpha$, alors on ne rejette pas H_0.

6.1 Tests paramétriques

On se place dans le cas de modèles gaussiens. On dispose d'un échantillon de la loi $\mathcal{N}(\mu, \sigma^2)$.

Tests sur la moyenne

On teste $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.

- Si la variance est connue : sous H_0, la VA $\overline{X_n}$ suit une loi $\mathcal{N}(\mu_0, \sigma^2/n)$, on va donc poser comme statistique $Z = \frac{\overline{X_n}-\mu_0}{\sqrt{\sigma^2/n}}$: sous H_0, $Z \sim \mathcal{N}(0,1)$, et en posant q le quantile d'ordre $1-\alpha/2$ de $\mathcal{N}(0,1)$, on a $\mathbb{P}(|Z| \leqslant q) = 1-\alpha$. La région de rejet associée est donc $]-\infty, -q[\cup]q, +\infty[$.
- Si la variance est inconnue : sous H_0, $\overline{X_n} \sim \mathcal{N}(\mu_0, \sigma^2/n)$ est indépendante de $\hat{S_n^2}$ qui vérifie : $(n-1)\hat{S_n^2}/\sigma^2 \sim \chi^2(n-1)$ (d'après Cochran), donc la statistique $T = \frac{\overline{X_n}-\mu_0}{\sqrt{\hat{S_n^2}/n}}$ suit une loi de $T(n-1)$. On prend alors q le quantile d'ordre $1-\alpha/2$ de la loi $T(n-1)$ et on procède comme précédemment.

Tests sur la variance

On teste $H_0 : \sigma^2 = \sigma_0^2$ contre $H_1 : \sigma^2 \neq \sigma_0^2$.

- Si la moyenne est connue : sous H_0, la statistique $V = \frac{nS_n^2}{\sigma_0^2} \sim \chi^2(n)$. La zone de rejet est alors $[0, q_{\alpha/2}[\cup]q_{1-\alpha/2}, +\infty[$, où q_x est le quantile d'ordre x de $\chi^2(n)$.
- Si la moyenne est inconnue : dans la statistique de test, on remplace m par $\overline{X_n}$.

6.2 Tests du χ^2

On se place dans le cas de VA discrètes, on effectue ici des tests asymptotiques. On considère des VA $(X_1, ..., X_n)$ i.i.d., de probabilité P, à valeurs dans un ensemble partitionné par $O_1, ..., O_m$. On pose :

- $N_k(n) = \sum_{i=1}^n \mathbf{1}_{X_i \in O_k}$, $N(n) = (N_1(n), ..., N_m(n))$.
- $p_k = \mathbb{P}(X_i \in O_k)$.
- $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_k}$ (loi empirique).
- $D(P_n, P) = \sum_{i=1}^m \frac{(N_k(n)-np_k)^2}{np_k}$.

On a alors le théorème suivant : $D(P_n, P) \xrightarrow{\mathcal{L}} \chi^2(m-1)$ (qui se démontre à l'aide du théorème de Cochran).

Remarque. Pour le test du chi-deux, il faut que les np_k soient $>> 5$. Si ce n'est pas le cas, on fusionne les O_i d'effectif faible.

Test d'adéquation

- On teste $H_0 : P = P^0$ contre $P \neq P^0$.
- Sous H_0, on a $D(P_n, P^0) \xrightarrow{\mathcal{L}} \chi^2(m-1)$ et sinon, $D(P_n, P^0) \to +\infty$ p.s.
- On pose alors $T = D(P_n, P^0)$ la statistique de test.
- On choisit la valeur critique s telle que $\mathbb{P}(K \geqslant s) = \alpha$ (quantile d'ordre $1-\alpha$), où $K \sim \chi^2(m-1)$, et la zone de rejet est alors $[s, +\infty[$.

Test d'adéquation à une famille de lois

Soit $X_1, ..., X_n$ un échantillon de X à valeurs dans $\{1, ..., m\}$, qui suit une loi P_θ avec $\theta \in \mathbb{R}^d$ inconnue. Soit $\hat{\theta}$ un estimateur de θ tel que $p_k(\hat{\theta}) \to p_k(\theta)$ (en général, on prend l'EMV). Ce test consiste à remplacer le paramètre θ par un estimateur.

- On teste $H_0 : P \in (P_\theta)_\theta$ contre $H_1 : P \notin (P_\theta)_\theta$.
- Sous H_0, on a $D(P_n - P(\hat{\theta})) \xrightarrow{\mathcal{L}} \chi^2(m-d-1)$. On pose alors $D(P_n - P(\hat{\theta}))$ la statistique de test, puis on raisonne comme précedemment.

Test d'indépendance

On considère un échantillon $((X_1, Y_1), ...(X_n, Y_n))$ d'un couple (X, Y) avec X à valeurs dans $\{a_1, ..., a_r\}$ et Y à valeurs dans $\{b_1, ..., b_t\}$.

- On teste $H_0 : X \perp\!\!\!\perp Y$ contre H_1 : X et Y ne sont pas indépendants. On pose $N_{i,j}$ le nombre de couples (a_i, b_j) observés, $N_{i,.} = \sum_{j=1}^t N_{i,j}$ et $N_{.,j} = \sum_{i=1}^r N_{i,j}$.
- Sous H_0, on a $T := n\sum_{i=1}^r \sum_{j=1}^t \frac{(N_{i,j} - \frac{N_{i,.}N_{.,j}}{n})^2}{N_{i,.}N_{.,j}} \xrightarrow{\mathcal{L}} \chi^2((r-1)(t-1))$. On pose alors T la statistique de test, puis on raisonne comme précédemment.

6.3 Tests de Kolmogorov-Smirnov

On se place dans le cas de VA continues, on effectue ici des tests asymptotiques.

- *Thm.* Soit $(X_1, ..., X_n)$ un échantillon issu de X. On note F la FDR de X et F_n la FDR empirique. Si F est continue, alors la loi de la statistique $D(F_n, F) := \sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$ ne dépend pas de F.
- *Thm.* Sous les mêmes hypothèses, la VA $\sqrt{n}D(F_n, F)$ converge en loi vers une loi limite qui ne dépend pas de F, qu'on appelle loi de Kolmogorov-Smirnov, et dont la FDR est donnée par : $\forall x > 0$,

$$F_{KS}(x) = 1 - 2\sum_{k\geqslant 1}(-1)^{k+1}\exp(-2k^2x^2).$$

- *Remarque.* Pour calculer $D(F_n, F)$ en pratique, on a $D(F_n, F) = \max_{1 \leqslant i \leqslant n} \max \left(\left| \frac{i}{n} - F(X_{(i)}) \right|, \left| \frac{i-1}{n} - F(X_{(i)}) \right| \right)$.

Test d'adéquation

- On teste $H_0 : F = F_0$ contre $H_1 : F \neq F_0$.
- Sous H_0, on a $T := \sqrt{n} D(F_n, F) \xrightarrow{\mathcal{L}} \mu_{KS}$, et sinon, $T \to +\infty$ p.s. On pose alors T la statistique de test.
- On choisit la valeur critique s telle que $\mathbb{P}(KS \geqslant s) = \alpha$ où $KS \sim \mu_{KS}$, et la zone de rejet est alors $[s, \infty[$.

Test d'homogénéité

On considère un échantillon $(X_1, ..., X_n)$ de f.d.r. F, et un échantillon $(Y_1, ..., Y_m)$ de f.d.r. G.

- On veut tester $H_0 : F = G$ contre $H_1 : F \neq G$. On pose F_n et G_m les f.d.r. empiriques, et $D_{m,n} = \sup_{t \in \mathbb{R}} |F_n(t) - G_m(t)|$.
- Sous H_0, on a $T := \sqrt{\frac{nm}{n+m}} D_{m,n} \xrightarrow{\mathcal{L}} \mu_{KS}$. On prend alors T la statistique de test, puis on raisonne comme précédemment.

7 Quelques contre-exemples

1. **Suite de VA qui converge en loi mais dont les densités ne convergent pas (cf [Hau]) :**
   On considère X_n de FDR $F_n : x \mapsto x - \frac{\sin(2\pi n x)}{2\pi n}$ sur $[0, 1]$. Alors $(F_{X_n})_n$ converge simplement vers la fonction $x \mapsto x$, donc $X_n \xrightarrow{\mathcal{L}} \mathcal{U}(0, 1)$, or f_{X_n} oscille de plus en plus donc n'a pas de limite.

2. **Suites de VA qui convergent en loi mais dont la somme ne converge pas en loi vers la somme des limites :**
   On considère $X \sim \mathcal{N}(0, 1)$ et, pour tout $n \geqslant 1$, $X_n = -X$. Alors $X_n \xrightarrow{\mathcal{L}} X$, mais $X_n - X = -2X$ ne tend pas vers 0.

3. **CV en proba mais pas p.s (cf [Hau]) :**
   On considère la suite d'intervalles $[0, 1]$, $[0, 1/2]$, $[1/2, 1]$, $[0, 1/3]$, $[1/3, 2/3]$, $[2/3, 1]$, ... et (X_n) la suite des fonctions indicatrices sur ces intervalles. Alors $X_n \xrightarrow{\mathbb{P}} 0$, mais pour tout $w \in [0, 1]$, $X_n(w) = 0$ pour une infinité de n, donc il n'y a pas convergene p.s.

4. **CV en proba mais pas L^1 :**
   On considère X_n définie sur $[0, 1]$ par $X_n(w) = n 1_{]0,1/n[}(w)$, alors $X_n \xrightarrow{ps} 0$, donc $X_n \xrightarrow{\mathbb{P}} 0$, mais $\mathbb{E}[X_n] = 1$ ne tend pas vers 0.

5. **CV L^1 mais pas L^2 :**
   On considère X_n définie sur $[0, 1]$ par $X_n(w) = \sqrt{n} 1_{]0,1/n[}(w)$, alors $\mathbb{E}[X_n] \underset{n \to \infty}{\longrightarrow} 0$, mais $\mathbb{E}[X_n^2] = 1$ ne tend pas vers 0.

6. **VA gaussiennes dont la somme n'est pas gaussienne :**
   On considère $X \sim \mathcal{N}(0, 1)$ et ε une variable de loi de Rademacher. On pose $Y = \varepsilon X$. Alors on a $Y \sim \mathcal{N}(0, 1)$, mais $X + Y$ n'est pas gaussienne car il y a un atome en 0, ce qui n'est pas possible pour une variable gaussienne.

7. **VA non corrélées mais non indépendantes :**
   On considère X une variable aléatoire de loi uniforme sur $\{-1, 0, 1\}$, et on pose $Y = X^2$. Alors on peut montrer que X, Y sont non corrélées, mais elles ne sont pas indépendantes.

References

[App] Appel. *Probabilités pour les non-probabilistes.*

[Ben] Benaïm. *Promenade aléatoire.*

[CR] Chabanol-Ruch. *Probabilités et statstiques pour l'épreuve de modélisation à l'agrégation de mathématiques.*

[GK] Garet-Kurtzmann. *De l'intégration aux probabilités.*

[Hau] Hauchecorne. *Contre-exemples en mathématiques.*

[RS] Rivoirard-Stoltz. *Statistique en action.*