Université Savoie Mont-Blanc **L3 MIASHS Éco**

MATH602 - Formulaire d'échantillonage

1) Convergence des variables aléatoires, et propriété des échantillons

Définition 1. Soient $(X_n)_n$ et X des v.a. définies sur un même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

1. $X_n \to X$ *presque partout* si pour presque tout $x \in \Omega$, $X_n(x) \to X(x)$.
2. $X_n \to X$ *en probabilité* si pour tout $\varepsilon > 0$, $P(|X_n - X| \geq \varepsilon) \to 0$.

Loi forte des grands nombres Soit $(X_n)_n$ une suite de v.a. sur Ω. On suppose :

1. Les (X_i) sont indépendantes (ind.)
2. Les (X_i) ont la même loi (id. dist.)
3. Les (X_i) sont intégrables et $E(X_1) = \mu$

Alors $\displaystyle\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to+\infty}{\longrightarrow} \mu$ presque partout sur Ω.

Définition 2. Soient (X_n) et X des v.a. $X_n \to X$ *en loi* si l'une de ces propriétés est satisfaite :

1. Pour presque tout $x \in \mathbb{R}$, $F_{X_n}(x) \underset{n\to+\infty}{\longrightarrow} F_X(x)$.
2. Pour toute fonction continue bornée f, $E(f(X_n)) \underset{n\to+\infty}{\longrightarrow} E(f(X))$.
3. Pour tout $t \in \mathbb{R}$, $E(e^{itX_n}) \underset{n\to+\infty}{\longrightarrow} E(e^{itX})$.

Théorème central-limit Soit (X_n) une suite de v.a. et $Z \sim \mathcal{N}(0,1)$. On suppose :

1. Les (X_i) sont indépendantes (ind.)
2. Les (X_i) ont la même loi (id. dist.)
3. Les (X_i) sont de carrés intégrables, et $E(X_1) = \mu$, $Var(X_1) = \sigma^2$.

Alors $\displaystyle\frac{1}{\sqrt{n}\sigma}\Big(\sum_{i=1}^{n} X_i - n\mu\Big) \underset{n\to+\infty}{\longrightarrow} Z$ en loi, ou encore $\displaystyle\frac{\sqrt{n}}{\sigma}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\Big) \underset{n\to+\infty}{\longrightarrow} Z$ en loi.

Intervalles de confiance pour X quand μ et σ^2 sont connues Soit X de carré intégrable,

$$P\Big(\mu - R \quad \leq \quad X \quad \leq \quad \mu + R\Big) \geq 1 - \frac{\sigma^2}{R^2}.$$

Intervalles de confiance pour $\frac{1}{n}\sum X_i$ et $\sum X_i$, quand μ et σ^2 sont connues
Soit (X_n) une suite de v.a. ind. id. dist. de carré intégrable, et $\alpha \in]0,1[$ un risque. Alors

$$P\Big(\mu - \frac{z_{1-\alpha/2} \times \sigma}{\sqrt{n}} \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n} X_i \quad \leq \quad \mu + \frac{z_{1-\alpha/2} \times \sigma}{\sqrt{n}}\Big) = 1 - \alpha,$$

où $\mu = E(X_1)$ et $\sigma^2 = Var(X_1)$, et

$$P\Big(n\mu - z_{1-\alpha/2}\sqrt{n} \times \sigma \quad \leq \quad \sum_{i=1}^{n} X_i \quad \leq \quad n\mu + z_{1-\alpha/2}\sqrt{n} \times \sigma\Big) = 1 - \alpha.$$

Quantiles remarquables de la loi normale Soient $Z \sim \mathcal{N}(0,1)$ et $\alpha \in]0,1[$ un risque.
On définit $z_{1-\alpha/2}$ par $P(Z \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]) = 1 - \alpha$ (la certitude).

α	1	0,8	0,5	0,32	0,2	0,1	0,05	0,0455	0,01	0,0027	0,001	0,000063
$1-\alpha$	0	0,2	0,5	0,68	0,8	0,9	0,95	0,9545	0,99	0,9973	0,999	0,999937
$z_{1-\alpha/2}$	0	0,25	0,67	1	1,28	1,65	1,96	2	2,58	3	3,29	4

2) Estimation par échantillonage

$(X_n)_n$: suite de v.a.ind.id.dist, de même loi que $X_\theta \sim \mathcal{L}(\theta)$, où $\mathcal{L}(\theta)$ est une loi et θ est un paramètre inconnu de $\mathcal{L}$. On suppose aussi que X_θ est de carré intégrable : $E(X_\theta)$ et $Var(X_\theta)$ existent.

Définition 3. Un *estimateur* T de θ est une suite de fonctions $(T_n)_n$ sur $\mathbb{R}^n$.
- Pour tout $n \in \mathbb{N}$, $T_n(X_1, \ldots, X_n)$ est une variable aléatoire "censée" converger vers θ.
- T est *sans biais* si $\forall n \in \mathbb{N}$, $E(T_n(X_1, \ldots, X_n)) = \theta$.
- T est *asymptotiquement sans biais* si $\lim_{n\to+\infty} E(T_n(X_1, \ldots, X_n)) = \theta$.
- T est *consistant* si $T_n(X_1, \ldots, X_n)$ converge en probabilité vers θ.
- *L'erreur quadratique moyenne* est la suite $E((T_n(X_1, \ldots, X_n) - \theta)^2)_n$.
- Pour un *échantillon* $(x_1, \ldots, x_n) \in \mathbb{R}^n$, le nombre $T_n(x_1, \ldots, x_n)$ est *une estimation* de θ.
- La *vraisemblance* L d'un échantillon $(x_1, \ldots, x_n)$, selon la loi parente X_θ, est :

$$L(x_1, \ldots, x_n, \theta) = \prod_{k=1}^{n} P(X_\theta = x_k) \text{ (cas discret)} \quad \text{ou} \quad L(x_1, \ldots, x_n, \theta) = \prod_{k=1}^{n} f_\theta(x_k) \text{ (cas continu)}$$

Définition 4. Soit $(x_1, \ldots, x_n)$ un échantillon. Les estimateurs classiques :
- *Moyenne empirique* $\mathcal{M}(x_1, \ldots, x_n) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} x_k$. estimateur de $E(X_\theta)$.
- *Variance empirique* $\mathcal{V}(x_1, \ldots, x_n) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(x_k - \mathcal{M}(x_1, \ldots, x_n))^2$ estimateur de $Var(X_\theta)$.
- *Variance sans biais* $\mathcal{V}^*(x_1, \ldots, x_n) = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(x_k - \mathcal{M}(x_1, \ldots, x_n))^2$ estimateur de $Var(X_\theta)$.
- *L'estimateur du maximum de vraisemblance* est $\widehat{\theta}$ tel que $L(x_1, \ldots, x_n, \widehat{\theta})$ est maximale.
- *L'estimateur des moments* est la solution θ_0 de l'équation : $\mathcal{M}(x_1, \ldots, x_n) = E(X_\theta)$.

Intervalles de confiance pour $E(X_\theta)$ et $Var(X_\theta)$, quand $\mathcal{M}$ et $\mathcal{V}^*$ sont connues
Soit (X_n) une suite de v.a. ind. id. dist. de lois Gaussiennes $\mathcal{N}(\mu, \sigma^2)$, et $\alpha \in]0,1[$ un risque. Alors

$$P\Big(\mathcal{M} - t^{n-1}_{1-\alpha/2}\frac{\sqrt{\mathcal{V}^*}}{\sqrt{n}} \quad\leq\quad \mu \quad\leq\quad \mathcal{M} + t^{n-1}_{1-\alpha/2}\frac{\sqrt{\mathcal{V}*}}{\sqrt{n}}\Big) = 1 - \alpha,$$

et

$$P\Big(\frac{(n-1)\mathcal{V}^*}{k^{n-1}_{1-\alpha/2}} \quad\leq\quad \sigma^2 \quad\leq\quad \frac{(n-1)\mathcal{V}^*}{k^{n-1}_{\alpha/2}}\Big) = 1 - \alpha.$$

Quantiles lois de Student et χ^2 Soient $T \sim St(n-1)$, $K \sim \chi^2(n-1)$ et $\alpha \in]0,1[$ un risque.
On définit $t^{n-1}_{1-\alpha/2}$ par $P(T \in [-t^{n-1}_{1-\alpha/2}, t^{n-1}_{1-\alpha/2}]) = 1 - \alpha$ (quantile d'ordre $1 - \alpha/2$ de $St(n-1)$).
On définit $k^{n-1}_{\alpha/2}$ et $k^{n-1}_{1-\alpha/2}$ par $P(K \in [k^{n-1}_{\alpha/2}, k^{n-1}_{1-\alpha/2}]) = 1 - \alpha$ (quantiles $\alpha/2$ et $1 - \alpha/2$ de $\chi^2(n-1)$).

Risque	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,001$	$\alpha = 0,05$	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,01$
n	$t^{n-1}_{0,975}$	$t^{n-1}_{0,995}$	$t^{n-1}_{0,9995}$	$k^{n-1}_{0,025}$	$k^{n-1}_{0,975}$	$k^{n-1}_{0,005}$	$k^{n-1}_{0,995}$
2	$12,71$	$63,66$	636	0	$5,02$	0	$7,88$
3	$4,3$	$9,93$	$31,6$	$0,05$	$7,38$	$0,01$	$10,6$
5	$2,78$	$4,6$	$8,61$	$0,48$	$11,1$	$0,21$	$14,9$
8	$2,36$	$3,5$	$5,41$	$1,69$	16	$0,99$	$20,3$
10	$2,26$	$3,25$	$4,78$	$2,7$	19	$1,73$	$23,6$
20	$2,09$	$2,86$	$3,88$	$8,8$	$32,8$	$6,84$	$38,6$
30	$2,04$	$2,75$	$3,66$	16	$45,7$	$13,12$	$52,3$
50	$2,01$	$2,68$	$3,5$	$32,4$	$71,4$	28	$79,5$
∞	$1,96$	$2,58$	$3,29$	$n - 2\sqrt{n}$	$n + 2\sqrt{n}$	$n - 2,6\sqrt{n}$	$n + 2,6\sqrt{n}$