Correction du TD 2

Exercice 1

Montrons que $\frac{1}{n}\sum_{i=1}^{n} X_i$ converge en probabilité vers m, c'est-à-dire que :

$$\lim_{n \to +\infty} P(|X_n - m| < \varepsilon) = 1, \forall\ \varepsilon > 0$$

$$P(|X_n - m| < \varepsilon) = P(-\varepsilon < X_n - m < \varepsilon) = P\Big(-\frac{\sqrt{n}}{\sigma}\varepsilon < \frac{X_n - m}{\frac{\sigma}{\sqrt{n}}} < \frac{\sqrt{n}}{\sigma}\varepsilon\Big)$$

Comme on sait que les X_i suivent une loi normale $N(m, \sigma^2)$, on en déduit que la variable $Z_n = \frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}}$ suit une loi normale N(0,1), soit :

$$\lim_{n \to +\infty} P(|X_n - m| < \varepsilon) = \lim_{n \to +\infty} P\Big(-\frac{\sqrt{n}}{\sigma}\varepsilon < Z_n < \frac{\sqrt{n}}{\sigma}\varepsilon\Big) = P(-\infty < Z_n < +\infty) = 1$$

NB : La loi des grands nombres permet d'obtenir ce résultat sans faire avoir l'hypothèse d'une loi normale.

Exercice 2

Rappel :Inégalité de Chebyshev
Soit X une variable aléatoire réelle, discrète ou continue de variance V(X), alors

$$\forall\ \varepsilon > 0, P\Big(|X - E(X)|) > \varepsilon\Big) \leq \frac{V(X)}{\varepsilon^2}$$

On doit montrer $\lim_{n \to +\infty} P(|Y_n - Y|) > \varepsilon) = 0\ \forall\ \varepsilon > 0$, tout en utilisant l'inégalité de Chebyshev.

On a $Y_n - Y = Z_n$, avec $E(Z_n) = 0$, donc

$$P(|Y_n - Y|) > \varepsilon) = P(|Z_n|) > \varepsilon) = P(|Z_n - E(Z_n)|) > \varepsilon)$$

D'après l'inégalité de Chebyshev, $\forall\ \varepsilon > 0, P(|Z_n - E(Z_n)|) > \varepsilon) \leq \frac{V(Z_n)}{\varepsilon^2}$

$$\lim_{n \to +\infty} P(|Y_n - Y|) > \varepsilon) = \lim_{n \to +\infty} P(|Z_n - E(Z_n)|) > \varepsilon) \leq \lim_{n \to +\infty} \frac{V(Z_n)}{\varepsilon^2} = \lim_{n \to +\infty} \frac{1/n}{\varepsilon^2}$$

$$0 \leq \lim_{n \to +\infty} P(|Y_n - Y|) > \varepsilon) \leq 0 \Rightarrow \lim_{n \to +\infty} P(|Y_n - Y|) > \varepsilon) = 0$$

Exercice 3

D'après l'inégalité de Chebychev, on a

$$\forall\ \varepsilon > 0, P(|X - E(X)| > \varepsilon) \leq \frac{V(X)}{\varepsilon^2} \Leftrightarrow P(|X - m| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Si on pose $\varepsilon = \lambda\sigma$ alors :

$$P(|X - m| > \lambda\sigma) \leq \frac{1}{\lambda^2} \Rightarrow 1 - P(|X - m| > \lambda\sigma) \geq 1 - \frac{1}{\lambda^2} \Rightarrow P(|X - m| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}$$

Exercice 4

Quelles sont les différences entre les convergences presque sûres, en loi, en probabilité et en moyenne quadratique ?

Soit (X_n) une suite de va. Soit F_n la fonction de répartition de X_n. Soit X une va de fonction de répartition F.

Convergence en probabilité On dit que (X_n) converge en probabilité vers X ($X_n \xrightarrow{P} X$ ou $p\lim_{n\to+\infty} X_n = X$) si

$$\forall\varepsilon > 0, \lim_{n\to+\infty} Pr(|X_n - X| > \varepsilon) = 0$$

La convergence en probabilité est utilisée pour exprimer la loi faible des grands nombres.

Loi faible des grands nombres (Khinchine) Soit (X_i) une suite de variables aléatoires réelles indépendantes et identiquement distribuées (iid) d'espérance finie, $E(X_i) = m$:

$$\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{Alors } \bar{X_n} \xrightarrow{p} m$$

Convergence en loi On dit que (X_n) converge en loi vers X ($X_n \xrightarrow{L} X$) si la suite de fonctions (F_n) converge, point par point, vers F :

$$\forall x, \lim_{n\to+\infty} |F_n(X_n) - F(X)| = 0$$

La convergence en probabilité implique la convergence en loi, mais pas l'inverse. La convergence en loi est une propriété de la fonction de répartition mais pas de la variable aléatoire elle-même. Celle-ci peut très bien ne pas converger.

La convergence en loi est utilisée pour exprimer le théorème central limite.

Théorème central limite Soit une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires indépendantes, de même loi, d'espérance m et de variance finie σ^2. Notons $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Alors

$$\sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right) \xrightarrow{L} \mathcal{N}(0,1)$$

Convergence presque sûre On dit que (X_n) converge presque sûrement vers X ($X_n \xrightarrow{ps} X$ si et seulement si

$$\forall \varepsilon > 0, \lim_{n \to +\infty} Pr(|X_i - X| > \varepsilon \ \forall i \geq n) = 0$$

La convergence presque sûre se distingue de la convergence en probabilité par le fait qu'elle est plus générale ($\forall i \geq n$). Intuitivement, elle implique qu'une fois que la série X_n s'approche de X, elle ne s'en écarte pas. On voit rapidement avec la définition (prenons i=n) que la convergence presque sûre implique la convergence en probabilité.

$$X_n \xrightarrow{ps} X \Rightarrow X_n \xrightarrow{P} X$$

La convergence presque sûre est utilisée pour exprimer la loi forte des grands nombres.

Convergence en moyenne quadratique On dit que (X_n) converge en moyenne quadratique vers X ($X_n \xrightarrow{L^2} X$) si

$$\lim_{n \to +\infty} E[||X_n - X||^2] \longrightarrow 0$$

Exercice 5

1. L'estimateur optimal pour p est la fréquence observée :

$$\hat{p} = \frac{1}{600} \sum_i X_i = F$$

Les observations faites conduisent à $f = \frac{78}{600} = 0,13$.

2. loi de cet estimateur

$$600F \rightsquigarrow Binomiale(600, p)$$

Rappels sur la loi Binomiale $\mathcal{B}(n,p)$:
- *On réalise n épreuves successives indépendantes où on observe à chaque fois la réalisation ou non d'un certain événement A et on note X le nombre de réalisation de A.*

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

- *Si $X_i \rightsquigarrow B(1,p)$, alors $\sum_i^n X_i \rightsquigarrow \mathcal{B}(n,p)$.*
- *Si $n > 30$ et $np < 15$, alors la loi $\mathcal{B}(n,p)$ converge en loi vers la loi $\mathcal{P}(np)$.*
- *Si $np(1-p) > 15$, la loi $\mathcal{B}(np)$ converge en loi vers la loi $\mathcal{N}(np, np(1-p))$.*

Nous pouvons utiliser l'approximation normale pour les calculs numériques, car $np(1-p) = 600 \times 0,13 \times (1 - 0,13) = 67,4 > 15$.

3. Intervalle bilatéral de confiance proche de 95 % pour p :

$$nF \rightsquigarrow N\left(np, np(1-p)\right)$$

$$F \rightsquigarrow N\left(p, \frac{p(1-p)}{n}\right) \Leftrightarrow \frac{F - p}{\sqrt{\frac{p(1-p)}{600}}} \rightsquigarrow N(0,1)$$

On lit dans la table de la loi normale $P(|U| < 1,96) = 0,95$ d'où :

$$p\left(\left|\frac{F-p}{\sqrt{\frac{p(1-p)}{600}}}\right| < 1,96\right) = 0,95$$

Pour approcher l'inégalité nous remplaçons p au dénominateur par son estimation F et nous obtenons :

$$-1,96 < \frac{F-p}{\sqrt{\frac{F(1-F)}{600}}} < 1,96$$

$$p\left(F - 1,96\sqrt{\frac{F(1-F)}{600}} < p < F + 1,96\sqrt{\frac{F(1-F)}{600}}\right) \simeq 0,95$$

$$f - 1,96\sqrt{\frac{f(1-f)}{600}} < p < f + 1,96\sqrt{\frac{f(1-f)}{600}}$$

Réalisation observée : $0,10309 < p < 0,15691$

Exercice 9

Nous désirons estimer m et prévoir le prochain tirage X_{n+1}. La variance est connue et égale à 4.

1. La moyenne empirique de l'échantillon est l'estimateur du maximum de vraisemblance de m. Elle suit une loi normale :

$$\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \rightsquigarrow \mathcal{N}\left(m, \frac{4}{n}\right)$$

2. La meilleure prévision de X_{n+1} est celle qui minimise l'erreur quadratique moyenne (MSE) :

$$MSE = E\Big[(X_{n+1} - \hat{x})^2\Big]$$

$$MSE = E\Big[(X_{n+1} - m + m - \hat{x})^2\Big]$$

$$MSE = E\Big[(X_{n+1} - m)^2 + 2(X_{n+1} - m)(m - \hat{x}) + (m - \hat{x})^2\Big]$$

$$MSE = E\Big[(X_{n+1} - m)^2\Big] + 2E\Big[(X_{n+1} - m)\Big]E\Big[(m - \hat{x})\Big] + E\Big[(m - \hat{x})^2\Big]$$

$$MSE = V(X_{n+1}) + 2\Big[(E(X_{n+1}) - m)\Big]E\Big[(m - \hat{x})\Big] + E\Big[(m - \hat{x})^2\Big]$$

$$MSE = V(X_{n+1}) + E\Big[(m - \hat{x})^2\Big]$$

Puisque X_{n+1} est indépendant des observations déjà faites et d'espérance m inconnue, la meilleure prévision de X_{n+1} est l'estimateur de m qui minimise le risque quadratique moyen $E\Big[(m - \hat{x})^2\Big]$, c'est-à-dire $\bar{X_n}$ (la moyenne empirique).

3. L'erreur de prévision $Z = X_{n+1} - \bar{X_n}$ suit une loi normale. L'espérance de Z est $E(Z) = m - m = 0$, la variance de Z est $V(Z) = V(X_{n+1}) + V(\bar{X_n})$ car X_{n+1} et $\bar{X_n}$ sont indépendantes. Les expressions équivalentes pour $V(Z)$ sont :

$$V(Z) = \sigma^2 + \frac{\sigma^2}{n} = 4 + \frac{4}{n} = 4\Big(1 + \frac{1}{n}\Big) = 4\frac{n+1}{n}$$

Nous en déduisons la loi de l'erreur de prévision Z :

$$Z = X_{n+1} - \bar{X_n} \rightsquigarrow \mathcal{N}\Big(0, 4\frac{n+1}{n}\Big)$$

4. Pour obtenir un intervalle de prévision pour X_{n+1}, on réduit la variable Z :

$$\frac{X_{n+1} - \bar{X_n}}{\sqrt{\sigma^2 \frac{n+1}{n}}} \rightsquigarrow \mathcal{N}(0,1)$$

On cherche donc l'intervalle $[-u_n, u_n]$ qui vérifie :

$$p\Bigg(- u_n < \frac{X_{n+1} - \bar{X_n}}{\sqrt{\sigma^2 \frac{n+1}{n}}} < u_n\Bigg) = 0,80$$

D'après les tables statistiques, on a :

$$p\Bigg(- 1,2816 < \frac{X_{n+1} - \bar{X_n}}{\sqrt{\sigma^2 \frac{n+1}{n}}} < 1,2816\Bigg) = 0,80$$

On peut le vérifier aussi avec la table de la loi de Student (qui tend vers l'infini vers la loi normale centrée réduite).
On résout les inégalités et on obtient :

$$p\Bigg(\bar{X_n} - 1,2816\sqrt{\sigma^2\frac{n+1}{n}} < X_{n+1} < \bar{X_n} + 1,2816\sqrt{\sigma^2\frac{n+1}{n}}\Bigg) = 0,80$$

$$p\Bigg(\bar{X_n} - 2,616 < X_{n+1} < \bar{X_n} + 2,616\Bigg) = 0,80$$