Méthodes à retenir :

-

I. Applications directes du cours

Exercice 1 ☆☆ *Sondage*

Une population de personnes présente une propriété donnée avec une proportion inconnue $p \in]0,1[$.
On choisit un échantillon de n personnes et l'on pose $X_i = 1$ si le i-ème individu présente la propriété étudiée, 0 sinon.
On considère que les variables aléatoires X_i ainsi définies sont indépendantes et suivent toute une loi de Bernoulli de paramètre p.
a) Quelle est la loi suivie par

$$S_n = X_1 + \cdots + X_n\,?$$

b) Déterminer espérance et variance de S_n/n.
c) Soit $\varepsilon > 0$. Etablir

$$\mathbf{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leqslant \frac{1}{4n\varepsilon^2}$$

d) Pour $\varepsilon = 0,05$, quelle valeur de n choisir pour que S_n/n soit voisin de p à ε près avec une probabilité supérieure à 95 % ?

II. Exercices

Exercice 2 ☆☆

On considère $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes suivant une loi de Bernoulli de paramètre $p \in]0,1[$. On pose $Y_n = X_n + X_{n+1}$ et $S_n = \dfrac{Y_1 + \cdots + Y_n}{n}$.

Calculer $\lim\limits_{n\to+\infty} \mathbb{P}(S_n \geqslant 3p)$.

Exercice 3 ☆☆

Soit $\lambda > 0$. On note $(X_1, \ldots, X_n$ n v.a.i..d. de loi $\mathcal{P}(\lambda)$. On pose $S_n = \displaystyle\sum_{i=1}^{n} X_i$ et $T_n = \dfrac{S_n}{\sqrt{n\lambda}}$

1. Déterminer le loi de S_n, son espérance et sa variance.
2. En déduire l'espérance et la variance de T_n.
3. Pour tout $\varepsilon > 0$, montrer qu'il existe c_0 tel que :

$$\forall c \geq c_0, \forall n \in \mathbb{N}^*,\ \mathbb{P}[|T_n| \geq c] \leq \varepsilon$$

III. Exercices avancés

Exercice 4 ☆☆☆ *Modèle d'assurance*

Des sinistres surviennent sur une année avec une probabilité $p \in]0,1[$. Chaque sinistre à un coût fixe de C pour l'assureur. Chaque client paye une prime annuelle d'assurance π.

1. Pourquoi peut-on modéliser par une famille $(X_i)_{1 \leq i \leq N}$ de N variables aléatoires de loi $b(p)$ les N clients d'un assureur ?

2. Quelle est l'espérance de S le solde annuel des sinistres pour l'assureur ?

3. Quelle est la variance de S ?

4. Donner une valeur de π la prime annuelle d'un assuré pour que l'assureur soit rentable une année avec une probabilité supérieure à 99%.

Exercice 5 ☆☆☆ *estimateurs statistiques*

Soient $X_1, \ldots, X_n$ des variables mutuellement indépendantes suivant une même loi d'espérance m et de variance σ^2.

On pose $\bar{X}_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i$ et

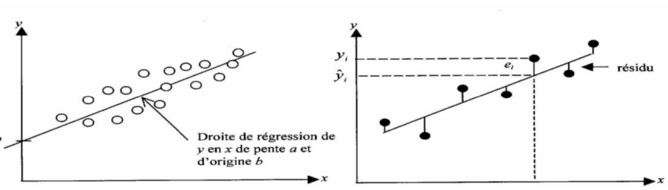
$$\bar{V}_n = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

a) Calculer espérance et variance de $\bar{X}_n$.

b) Calculer l'espérance de $\bar{V}_n$.

Exercice 6 ☆☆☆☆ *Modélisation par moindres carrés*

$$\sum_{i=0}^{n} \varepsilon_i^2 = \sum_{i=0}^{n} (y_i - (ax_i + b))^2.$$



Etant donné un n échantillon de mesures physiques $((x_i, y_i))_{1 \leq i \leq n}$, on cherche à savoir si cet échantillon peut-être assimilé à des réalisations aléatoires de couples $((x_i, ax_i + b))_{1 \leq i \leq n}$, pour des paramètres a et b de régression linéaire, selon une relation linéaire de la forme

$$y = a\ x + b$$

On note $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ et $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, et $U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathfrak{M}_{n,1}(\mathbb{R})$

1. Par définition, la droite de régression linéaire associée à ces mesures est la droite d'équation $y = \hat{a}x + \hat{b}$, où $(\hat{a}, \hat{b}) \in \mathbb{R}^2$ telle que $\delta_n(\hat{a}, \hat{b}) = \sum_{i=1}^{n} |y_i - \hat{a}\ x_i - \hat{b}|^2$ réalise le minimum de la fonction $(a, b) \mapsto \sum_{i=1}^{n} |y_i - a\ x_i - b|^2$.
2. Justifier que résoudre ce problème revient à minimiser une quantité de la forme
$$f(V) = \|MV - Y\|_2^2 \qquad (\star)$$
avec $M = [X, U] \in \mathfrak{M}_{n,2}(\mathbb{R})$ et $V = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathfrak{M}_{2,1}(\mathbb{R})$, et $Y \in \mathfrak{M}_{n,1}(\mathbb{R})$ à préciser.
3. On note $\hat{Y}$ le projeté orthogonal de Y sur le sous-espace vectoriel $\mathrm{Im}(M)$.
   (a) Pourquoi a-t-on $\mathrm{Im}(M) = \mathrm{Vect}(X, U)$?
   (b) Justifier que ce maximum est atteint et est unique.
   on utilisera la formule de projection sur un sous-espace vectoriel de dimension finie, muni d'une base orthonormée
   (c) Pourquoi a-t-on $\hat{Y} - Y \perp MV,\ \forall V \in \mathfrak{M}_{2,1}(\mathbb{R})$?
   (d) On pose $\hat{V}$ tel que $M\hat{V} = \hat{Y}$. En déduire que : $\forall\ V \in \mathfrak{M}_{2,1}(\mathbb{R}),\ V^T M^T (Y - M\hat{V}) = 0$
   (e) En déduire que : $M^T(Y - M\hat{V}) = 0$
   (f) Montrer que $\det(M^T M) = n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i^2)^2$
   (g) Etablir que $\hat{V} = (M^T M)^{-1}\ M^T\ Y$.
4. On rappelle que l'espérance de X (resp. Y) est $\mathbb{E}(X) = \dfrac{1}{n} \sum_{i=1}^{n} x_i$ (resp. $\mathbb{E}(Y) = \dfrac{1}{n} \sum_{i=1}^{n} y_i$).
   On rappelle que la variance de X (resp. Y) est $\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2$
   (resp. $\mathbb{V}(Y) = \mathbb{E}(Y - \mathbb{E}(Y))^2 = \dfrac{1}{n} \sum_{i=1}^{n} (y_i - \mathbb{E}(Y))^2$).
   On appelle covariance de X et Y le nombre : $\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \mathbb{E}(X))(y_i - \mathbb{E}(Y))$
   On vérifie que :
   $$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \left(\begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots x_n & \end{pmatrix} Y$$
   $$= \frac{1}{n^2 \mathbb{V}(X)} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}.$$
   En déduire que l'équation de la droite des moindres carrées est obtenue pour les paramètres : $\hat{b} = \mathbb{E}(Y) - \hat{a}\ \mathbb{E}(Y)$ et $\hat{a} = \dfrac{\mathrm{Cov}(X, Y)}{\mathbb{V}(X)}$.
5. La droite des moindres carrées et obtenue pour des paramètres $\hat{a}$ et $\hat{b}$ et a pour équation cartésienne
$$y = \hat{a}x + \hat{b}$$
Interpréter l'alignement des points sur cette droites dans le cas d'une covariance nulle.

Notes

[1] correction : a)S_n soit une loi de Bernoulli de paramètres n et p.

b) Puisque S_n suit une loi de Bernoulli, $\mathrm{E}(S_n) = np$ et $\mathrm{V}(S_n) = np(1-p)$.

Par conséquent $\mathrm{E}(S_n/n) = p$ et $\mathrm{V}(S_n/n) = p(1-p)/n$

c) Par l'inégalité de Bienaymé-Tchebychev $P(|Sn/n - \mathrm{E}(Sn/n)| > \varepsilon) \leq \mathrm{V}(Sn/n)/\varepsilon^2$

et donc $P(|Sn/n - p| > \varepsilon) \leq p(1-p)/(n\varepsilon^2)$

Enfin, l'inégalité classique $p(1-p) \leq 1/4$ permet de conclure.

d) On choisit n de sorte que $1/(4n\varepsilon^2) \leq 0,05$

La valeur $n = 2000$ est convenable.

[3] correction : B-T, $1c^2 \leq 1/c_0^2 \leq \varepsilon$