

Estimations Chapitre 14

Corrections exercices types

Exercice 1. 14-1

On considère une var X d'espérance m et de variance σ^2 et un n -échantillon $(X_1, \dots, X_n)$ de la loi de X . En notant $\bar{X}_n$ la moyenne empirique associée à cet échantillon, calculons le risque quadratique de $\bar{X}_n$.

On sait que $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur de m car il est fonction du n -échantillon $(X_1, \dots, X_n)$ de la loi de X et indépendant de m dans sa formulation.

On a par linéarité de l'espérance que $E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{m}{n} \sum_{i=1}^n 1 = \frac{nm}{n} = m$.

Donc la moyenne empirique est un estimateur sans biais de m (car $b_m(\bar{X}_n) = E(\bar{X}_n) - m = 0$).

Puis par indépendance des variables X_i (car $(X_1, \dots, X_n)$ est un n -échantillon de la loi de X) et par propriété de la variance : $V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n^2} \sum_{i=1}^n 1 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

Le risque quadratique est donné par :

$$r_m(\bar{X}_n) = E((\bar{X}_n - m)^2) = E(\bar{X}_n^2) - 2mE(\bar{X}_n) + m^2 = E(\bar{X}_n^2) - 2mE(\bar{X}_n) + m^2 + E(\bar{X}_n)^2 - E(\bar{X}_n)^2 = E(\bar{X}_n^2) - E(\bar{X}_n)^2 + E(\bar{X}_n)^2 - 2mE(\bar{X}_n) + m^2 = V(\bar{X}_n) + (E(\bar{X}_n) - m)^2 = V(\bar{X}_n) + b_m(\bar{X}_n).$$

On sait que la moyenne empirique est un estimateur sans biais de m donc le risque quadratique se résume à la variance de $\bar{X}_n$.

$$\text{D'où } r_m(\bar{X}_n) = V(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Exercice 2. 14-2

Montrons que la moyenne empirique est un estimateur convergent de l'espérance m de X .

C'est à dire montrons que : $\forall \epsilon > 0, \lim_{n \rightarrow +\infty} P(|\bar{X}_n - m| > \epsilon) = 0$.

Pour cela, appliquons l'inégalité de Bienaymé-Tchebychev à $\bar{X}_n$, on peut car $\bar{X}_n$, admet un moment d'ordre 2 : $\forall \epsilon > 0, P(|\bar{X}_n - E(\bar{X}_n)| \geq \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2}$.

Or, $E(\bar{X}_n) = m$ (cf exo 1 à savoir refaire) et $V(\bar{X}_n) = \frac{\sigma^2}{n}$ (idem voir exo 1 à savoir refaire).

D'où l'inégalité de B-T donne :

$$\forall \epsilon > 0, P(|\bar{X}_n - m| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

On a $|\bar{X}_n - m| > \epsilon \subset |\bar{X}_n - m| \geq \epsilon$ d'où par croissance des probabilités et par positivité des probabilités :

$$\forall \epsilon > 0, 0 \leq P(|\bar{X}_n - m| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

De plus $\lim_{n \rightarrow +\infty} \frac{\sigma^2}{n\epsilon^2} = 0$ donc par théorème d'encadrement, on en déduit que $\lim_{n \rightarrow +\infty} P(|\bar{X}_n - m| > \epsilon) = 0$. Donc la moyenne empirique est un estimateur convergent de l'espérance m de X .

NB : pour prouver q'un estimateur de θ , T_n est convergent, deux cas se présentent :

- s'il est sans biais alors on lui appliquera l'inégalité de B-T
- sinon on appliquera l'inégalité de Markov à $[(T_n - \theta)^2]$.

Exercice 3. 14-3

Une var X suit une loi uniforme sur $[0, \theta]$, θ étant un réel strictement positif inconnu. On note F la fonction de répartition de X et f une densité de X .

On dispose d'un n -échantillon $(X_1, \dots, X_n)$ de la loi de X et on pose $Z_n = \max(X_1, \dots, X_n)$. Déterminons la fonction de répartition F_n ainsi qu'une densité de Z_n .

$$Z_n = \max(X_1, \dots, X_n). \text{ On a alors } \forall x \in \mathbb{R}, [Z_n \leq x] = \left[\bigcap_{i=1}^n X_i \leq x \right].$$

L'indépendance des X_i (car n -échantillon $(X_1, \dots, X_n)$ de la loi de X), implique l'égalité suivante :

$$\forall x \in \mathbb{R}, F_n(x) = P([Z_n \leq x]) = \prod_{i=1}^n P([X_i \leq x]) = (P([X \leq x]))^n = (F(x))^n \text{ (car } X_i \text{ et } X \text{ suivent la même loi)}.$$

$$\text{Donc } F_n(x) = P([Z_n \leq x]) = (F(x))^n = \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{x}{\theta}\right)^n & \text{si } x \in [0, \theta] \\ 1 & \text{si } x > \theta \end{cases}$$

Z_n est à densité car $F_n(x) = (F(x))^n$ est continue sur $\mathbb{R}$ par composée de fonctions continues sur $\mathbb{R}$ (F et la fonction puissance n) et C^1 sur $\mathbb{R}$ sauf peut-être en 0 et θ par composées.

Pour avoir une densité, on dérive F_n partout où cela est possible.

Une densité possible pour Z_n est donnée par :

$$f_n(x) = \begin{cases} n \frac{x^{n-1}}{\theta^n} & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$$

Montrons que Z_n est un estimateur asymptotiquement sans biais de θ , puis que c'est un estimateur convergent de θ .

Tout d'abord, Z_n est un estimateur de θ , car Z_n est fonction du n -échantillon $(X_1, \dots, X_n)$ de la loi de X et indépendant de θ dans sa formulation.

Il nous faut voir ensuite si $E(Z_n)$ existe à savoir si $\int_{-\infty}^{+\infty} t f_n(t) dt$ CVABS.

Or, cette intégrale se résume à la CV de $\int_0^\theta \frac{n}{\theta^n} t^n dt$ car $f_n(t) = 0$ sur $] -\infty, 0[$ et sur $] \theta, +\infty[$.

$$\text{Et } \int_0^\theta \frac{n}{\theta^n} t^n dt = \frac{n}{\theta^n} \left[\frac{1}{n+1} t^{n+1} \right]_0^\theta = \frac{n}{n+1} \theta.$$

Pour voir enfin si Z_n est un estimateur asymptotiquement sans biais de θ , il faut montrer que : $\lim_{n \rightarrow +\infty} E(Z_n) = \theta$.

$$\lim_{n \rightarrow +\infty} E(Z_n) = \lim_{n \rightarrow +\infty} \frac{n}{n+1} \theta = \theta \quad (\text{car } \frac{n}{n+1} \underset{n \rightarrow +\infty}{\sim} 1)$$

Par conséquent, Z_n est bien un estimateur asymptotiquement sans biais de θ .

Pour prouver que Z_n est un estimateur convergent de θ , il nous faut appliquer l'inégalité de Markov à $(Z_n - \theta)^2$. Pour ce faire, il faut donc prouver que $E(Z_n^2)$ existe puis calculer $V(Z_n)$.

On s'intéresse donc à la CVABS de $\int_{-\infty}^{+\infty} t^2 f_n(t) dt$ CVABS.

Or, cette intégrale se résume à la CV de $\int_0^\theta \frac{n}{\theta^n} t^{n+1} dt$ car $f_n(t) = 0$ sur $] -\infty, 0[$ et sur $] \theta, +\infty[$.

$$\text{Et } \int_0^\theta \frac{n}{\theta^n} t^{n+1} dt = \frac{n}{\theta^n} \left[\frac{1}{n+2} t^{n+2} \right]_0^\theta = \frac{n}{n+2} \theta^2.$$

$$\text{Enfin, } V(Z_n) = E(Z_n^2) - E(Z_n)^2 = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1} \theta \right)^2 = \frac{n}{(n+2)(n+1)^2} \theta^2.$$

L'inégalité de Markov appliquée à $(Z_n - \theta)^2$ (qui est possible car $(Z_n - \theta)^2$ est une variable positive et telle que $E((Z_n - \theta)^2) = V(Z_n) + (E(Z_n) - \theta)^2 = \frac{n}{(n+2)(n+1)^2} \theta^2 + \left(\frac{n}{n+1} - 1 \right)^2 \theta^2 = \frac{n}{(n+2)(n+1)^2} \theta^2 + \frac{1}{(n+1)^2} \theta^2$

$$\forall a > 0, P(|(Z_n - \theta)^2| \geq a) \leq \frac{E((Z_n - \theta)^2)}{a}.$$

Avec $a = \epsilon^2$, cela donne :

$$\forall \epsilon > 0, P(|(Z_n - \theta)^2| \geq \epsilon^2) \leq \frac{E((Z_n - \theta)^2)}{\epsilon^2}.$$

Or, $|Z_n - \theta|^2 \geq \epsilon^2 \iff |Z_n - \theta| \geq \epsilon$, d'où

$$\forall \epsilon > 0, P(|Z_n - \theta| \geq \epsilon) \leq \frac{n}{(n+2)(n+1)^2 \epsilon^2} \theta^2 + \frac{1}{(n+1)^2 \epsilon^2} \theta^2$$

On a $|Z_n - \theta| > \epsilon \iff |Z_n - \theta| \geq \epsilon$ d'où par croissance des probabilités et par positivité des probabilités :

$$\forall \epsilon > 0, 0 \leq P(|Z_n - \theta| > \epsilon) \leq \frac{n}{(n+2)(n+1)^2 \epsilon^2} \theta^2 + \frac{1}{(n+1)^2 \epsilon^2} \theta^2.$$

De plus $\lim_{n \rightarrow +\infty} \frac{n}{(n+2)(n+1)^2 \epsilon^2} \theta^2 + \frac{1}{(n+1)^2 \epsilon^2} \theta^2 = 0$ donc par théorème d'encadrement, on en déduit que

$$\lim_{n \rightarrow +\infty} P(|Z_n - \theta| > \epsilon) = 0.$$

Donc Z_n est bien un estimateur convergent de θ .

$$\text{NB : le risque quadratique de } Z_n \text{ est : } r_\theta(Z_n) = V(Z_n) + (E(Z_n) - \theta)^2 = \frac{n}{(n+2)(n+1)^2} \theta^2 + \frac{1}{(n+1)^2} \theta^2 \xrightarrow{n \rightarrow +\infty} 0$$

On retrouve la condition suffisante de CV d'un estimateur à savoir que son risque quadratique tende vers 0.

Exercice 4. 14-4 Intervalle de confiance grâce à l'inégalité de Bienaymé-Tchebychev

Soit X une var de Bernoulli de paramètre p , $0 < p < 1$ et soit un n -échantillon $(X_1, \dots, X_n)$ de la loi de X . En notant $\bar{X}_n$ la moyenne empirique associée à cet échantillon et $\alpha \in]0, 1]$, calculons l'espérance et la variance de $\bar{X}_n$.

$$\text{On a par linéarité de l'espérance que } E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{p}{n} \sum_{i=1}^n 1 = \frac{np}{n} = p.$$

Puis par indépendance des variables X_i (car $(X_1, \dots, X_n)$ est un n -échantillon de la loi de X) et par propriété de

$$\text{la variance : } V(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{p(1-p)}{n^2} \sum_{i=1}^n 1 = \frac{npq}{n^2} = \frac{p(1-p)}{n}.$$

Étudions sur $]0, 1[$, la fonction $h : p \mapsto p(1 - p)$. h est dérivable sur $]0, 1[$ en tant que produit de fonctions dérivables sur $]0, 1[$.

$h'(p) = 1 - 2p$, $h'(p) > 0$ pour $0 < p < 1/2$, donc h est strictement croissante sur $]0, 1/2[$ et strictement décroissante sur $]1/2, 1[$. Sa dérivée s'annule en changeant de signes en $1/2$ donc h présente un maximum en $1/2$ qui vaut $h(1/2) = 1/4$.

Donc $\forall p \in]0, 1[$, $p(1 - p) \leq 1/4$

On sait que $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur de p car il est fonction du n -échantillon $(X_1, \dots, X_n)$ de la loi de X et indépendant de p dans sa formulation.

Appliquons l'inégalité de B-T à $\bar{X}_n$ (on peut car $\bar{X}_n$ admet un moment d'ordre 2) :

$$\forall \epsilon > 0, P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{V(\bar{X}_n)}{\epsilon^2}$$

C'est à dire que :

$$\forall \epsilon > 0, P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2} \quad (*)$$

On cherche un intervalle de confiance de p au niveau de confiance de $1 - \alpha$.

Or, $P(|\bar{X}_n - p| \geq \epsilon) = 1 - P(|\bar{X}_n - p| \leq \epsilon)$.

Donc $(*)$ donne :

$$\forall \epsilon > 0, P(|\bar{X}_n - p| \leq \epsilon) \geq 1 - \frac{1}{4n\epsilon^2}.$$

On cherche ϵ tel que $\alpha = \frac{1}{4n\epsilon^2} \leftrightarrow \epsilon = \frac{1}{2\sqrt{n\alpha}}$

D'où l'intervalle de confiance de p au niveau de confiance de $1 - \alpha$:

$$\left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}} \right]$$

Exercice 5. 14-5

On admet que la durée de vie, exprimée en jours, d'un composant électronique suit une loi normale d'écart-type 70. Les durées de vie de 250 composants ont donné une moyenne de 450 jours. Donnons un intervalle de confiance à 99% de durée de vie d'un composant électronique.

Si on note, X_i la variable égale à la durée de vie du i ème composant et m la durée de vie moyenne d'un composant, on a :

$$n = 250, X_i \hookrightarrow \mathcal{N}(m, 70^2), \bar{X}_n = 450, t_\alpha = 2,58.$$

$$\text{On a alors } P(450 - 2,58 \times \frac{70}{\sqrt{250}} \leq m \leq 450 + 2,58 \times \frac{70}{\sqrt{250}}) = 0,99$$

D'où $m \in [438; 462]$ avec un risque inférieur à 1%.

Exercice 6. 14-6

On effectue un contrôle de fabrication sur des pièces usinées dont une proportion p est défectueuse. On en contrôle 200 et on en trouve 20 de défectueuses.

Donnons des intervalles de confiance de p avec un niveau de confiance de 95% puis de 99%.

On note X_i la var de Bernoulli égale à 1 si la i ème pièce est défectueuse, alors $X_i \hookrightarrow \mathcal{B}(p)$.

$$n = 200, \bar{X}_n = 20/200 = 1/10, t_\alpha = 1,96 \text{ ou } 2,58.$$

$$\text{On a alors } P(1/10 - t_\alpha \times \frac{1}{2\sqrt{200}} \leq p \leq 1/10 + t_\alpha \times \frac{1}{2\sqrt{200}}) \geq 1 - \alpha$$

D'où $p \in [0,0307; 0,1693]$ avec un risque inférieur à 5% et $p \in [0,00880, 0,1913]$ avec un risque inférieur à 1%.