

T.D. V - Estimation

I - Construction d'estimateurs

Solution de l'exercice 1.

1. D'après la définition, $Y_n = \sum_{i=1}^n X_i$ est une somme de n variables aléatoires indépendantes de même lois de Bernoulli de paramètre p . Ainsi, Y_n suit une loi binomiale, $Y_n \hookrightarrow \mathcal{B}(n, p)$.

2. Comme $Y_n \hookrightarrow \mathcal{B}(n, p)$, alors

$$\begin{aligned}\mathbf{E}[Y_n] &= np, \\ \mathbf{V}(Y_n) &= np(1-p).\end{aligned}$$

D'après la formule de Koenig-Huygens,

$$\begin{aligned}\mathbf{V}(Y_n) &= \mathbf{E}[Y_n^2] - \mathbf{E}[Y_n]^2 \\ \mathbf{E}[Y_n^2] &= \mathbf{V}(Y_n) + \mathbf{E}[Y_n]^2 \\ &= np(1-p) + (np)^2 \\ &= np(1-p+np) \\ &= np(1+(n-1)p).\end{aligned}$$

3. En utilisant la question précédente et la linéarité de l'espérance,

$$\begin{aligned}\mathbf{E}[Y_n^2] &= np(1+(n-1)p) \\ \mathbf{E}[n^2 \bar{X}_n^2] &= np(1+(n-1)p) \\ \mathbf{E}[\bar{X}_n^2] &= \frac{p(1+(n-1)p)}{n}.\end{aligned}$$

Ainsi, \bar{X}_n^2 est un estimateur sans biais de $\frac{p(1+(n-1)p)}{n}$ et n'est donc pas un estimateur sans biais de p^2 (sauf si $n=1$).

Pour obtenir un estimateur sans biais de p^2 , on utilise le fait que

$\mathbf{E}[X_n] = p$. Ainsi,

$$\begin{aligned}\mathbf{E}[\bar{X}_n^2] &= \frac{p}{n} + \frac{n-1}{n}p^2 \\ \mathbf{E}[\bar{X}_n^2] &= \frac{\mathbf{E}[\bar{X}_n]}{n} + \frac{n-1}{n}p^2 \\ \mathbf{E}[\bar{X}_n^2] - \frac{1}{n}\mathbf{E}[\bar{X}_n] &= \frac{n-1}{n}p^2 \\ \frac{n}{n-1}\mathbf{E}\left[\bar{X}_n^2 - \frac{1}{n}\bar{X}_n\right] &= p^2 \\ \mathbf{E}\left[\frac{n}{n-1}\bar{X}_n^2 - \frac{1}{n-1}\bar{X}_n\right] &= p^2.\end{aligned}$$

Ainsi, $\frac{n}{n-1}\bar{X}_n^2 - \frac{1}{n-1}\bar{X}_n$ est un estimateur sans biais de p^2 .

Remarque. Comme X_i est à valeurs dans $\{0, 1\}$, alors $X_i^2 = X_i$. Ainsi,

$$\begin{aligned}\frac{n}{n-1}\bar{X}_n^2 - \frac{1}{n-1}\bar{X}_n &= \frac{n}{n-1} \times \left(\frac{\sum_{i=1}^n X_i}{n}\right)^2 - \frac{\sum_{i=1}^n X_i}{(n-1)n} \\ &= \frac{\sum_{i=1}^n X_i^2 + \sum_{1 \leq i \neq j \leq n} X_i X_j - \sum_{i=1}^n X_i^2}{n(n-1)} \\ &= \frac{\sum_{1 \leq i \neq j \leq n} X_i X_j}{n(n-1)} \\ &= \frac{2 \sum_{1 \leq i < j \leq n} X_i X_j}{n(n-1)} \\ &= \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} X_i X_j.\end{aligned}$$

Ainsi, un estimateur sans biais de p^2 est $\binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} X_i X_j$ (et il est

aisé de vérifier directement que l'espérance de cet estimateur est bien p^2).

2^e méthode. En utilisant la variance empirique, $\bar{X}_n - \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur sans biais de $p - p(1-p) = p^2$. De plus, comme $X_i(\Omega) = \{0, 1\}$, alors $X_i^2 = X_i$ et

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \\ &= \frac{1}{n-1} \sum_{i=1}^n X_i - \frac{2}{n-1} \bar{X}_n \sum_{i=1}^n X_i + \frac{n}{n-1} \bar{X}_n^2 \\ &= \frac{n}{n-1} \bar{X}_n - \frac{2n}{n-1} \bar{X}_n^2 + \frac{n}{n-1} \bar{X}_n^2 \\ &= \frac{n}{n-1} \bar{X}_n - \frac{n}{n-1} \bar{X}_n^2 \\ \bar{X}_n - \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 &= \left(1 - \frac{n}{n-1}\right) \bar{X}_n + \frac{n}{n-1} \bar{X}_n^2 \\ &= \frac{n}{n-1} \bar{X}_n^2 - \frac{1}{n-1} \bar{X}_n. \end{aligned}$$

□

Solution de l'exercice 2.

1. Comme \bar{X}_n est une fonction de X_1, \dots, X_n , alors \bar{X}_n est bien un estimateur. De plus,

$$\mathbf{E}[\bar{X}_n] = \mathbf{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i] = \frac{1}{n} \sum_{i=1}^n m = m.$$

Ainsi, \bar{X}_n est un estimateur sans biais de m .

2. Comme les variables aléatoires X_1, \dots, X_n sont indépendantes, d'après la propriété de la variance,

$$\begin{aligned} \mathbf{V}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \mathbf{V}(X_i) = \sum_{i=1}^n \sigma^2 = n\sigma^2 \\ \mathbf{V}(\bar{X}_n) &= \mathbf{V}\left(\frac{1}{n} \bar{X}_n\right) = \frac{1}{n^2} \mathbf{V}(\bar{X}_n) = \frac{\sigma^2}{n}. \end{aligned}$$

3. En utilisant une identité remarquable, puis les propriétés de la somme,

$$\begin{aligned} ns_n &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}_n \sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}_n^2 \\ &= \sum_{i=1}^n X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2 \\ s_n &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}_n^2. \end{aligned}$$

Ainsi, en utilisant la linéarité de l'espérance et les questions précédentes,

$$\begin{aligned} \mathbf{E}[s_n] &= \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i^2] - \mathbf{E}[\bar{X}_n^2] \\ &= \frac{1}{n} \sum_{i=1}^n (\mathbf{V}(X_i) + \mathbf{E}[X_i]^2) - (\mathbf{V}(\bar{X}_n) + \mathbf{E}[\bar{X}_n]^2) \\ &= \frac{1}{n} \sum_{i=1}^n (\sigma^2 + m^2) - \left(\frac{\sigma^2}{n} + m^2\right) \\ &= \sigma^2 + m^2 - \left(\frac{\sigma^2}{n} + m^2\right) = \frac{n-1}{n} \sigma^2. \end{aligned}$$

4. D'après la question précédente, dès que $n \neq 1$, la variable aléatoire s_n n'est pas un estimateur sans biais de σ^2 .

5. S_n est une fonction de X_1, \dots, X_n donc c'est un estimateur. D'après la question précédente, $S_n = \frac{n}{n-1} s_n$. Ainsi,

$$\mathbf{E}[S_n] = \frac{n}{n-1} \mathbf{E}[s_n] = \frac{n}{n-1} \times \frac{n-1}{n} \sigma^2 = \sigma^2.$$

□

Solution de l'exercice 3.

1. On est dans le cadre du calcul de la fonction de répartition d'un maximum :

$$\begin{aligned}
 \mathbf{P}([M_n \leq i]) &= \mathbf{P}([\max \{X_1, \dots, X_n\} \leq i]) \\
 &= \mathbf{P}\left(\bigcap_{k=1}^n [X_k \leq i]\right) \\
 &= \prod_{k=1}^n \mathbf{P}([X_k \leq i]), \text{ en utilisant l'indépendance} \\
 &= \prod_{k=1}^n \mathbf{P}([X_1 \leq i]), \text{ car les tirages sont avec remise} \\
 &= \prod_{k=1}^n [X_1 \leq i]^n \\
 &= \left(\sum_{\ell=1}^i \mathbf{P}([X_1 = \ell])\right)^n, \text{ car } X_1 \text{ prend les valeurs } 1, \dots, N \\
 &= \left(\frac{i}{N}\right)^n, \text{ car les boules sont indistinguables}
 \end{aligned}$$

2. En utilisant la formule précédente,

$$\mathbf{P}([M_n = N]) = \mathbf{P}([M_n \leq N]) - \mathbf{P}([M_n \leq N-1]) = 1 - \left(\frac{N-1}{N}\right)^n.$$

Comme $-1 < \frac{N-1}{N} < 1$, alors $\lim_{n \rightarrow +\infty} \mathbf{P}([M_n = N]) = 1$.

3. On va effectuer un changement dans l'ordre des sommations :

$$\begin{aligned}
 \sum_{i=1}^N \mathbf{P}(Y \geq i) &= \sum_{i=1}^N \sum_{j=i}^N \mathbf{P}(Y = j) \\
 &= \sum_{1 \leq i \leq j \leq N} \mathbf{P}(Y = j) \\
 &= \sum_{j=1}^N \sum_{i=1}^j \mathbf{P}(Y = j) \\
 &= \sum_{j=1}^N j \mathbf{P}(Y = j) = \mathbf{E}[Y].
 \end{aligned}$$

4. En utilisant les questions précédentes, comme M_n est à valeurs dans $\llbracket 1, N \rrbracket$,

$$\begin{aligned}
 \mathbf{E}[M_n] &= \sum_{i=1}^N \mathbf{P}([M_n \geq i]) \\
 &= \sum_{i=1}^N (1 - \mathbf{P}([M_n \leq i-1])) \\
 &= \sum_{i=1}^N \left(1 - \left(\frac{i-1}{N}\right)^n\right) \\
 &= N - \sum_{i=1}^N \left(\frac{i-1}{N}\right)^n \\
 &= N - \sum_{i=0}^{N-1} \left(\frac{i}{N}\right)^n.
 \end{aligned}$$

5. D'après la question précédente, pour $k \in \llbracket 0, N-1 \rrbracket$,

$$\begin{aligned}
 \frac{0}{N} &\leq \frac{j}{N} \leq \frac{N-1}{N} \\
 0 &\leq \left(\frac{j}{N}\right)^n \leq \left(\frac{N-1}{N}\right)^n \\
 -\left(\frac{N-1}{N}\right)^n &\leq \left(\frac{j}{N}\right)^n \leq 0 \\
 -\sum_{j=0}^{N-1} \left(1 - \frac{1}{N}\right)^n &\leq \sum_{j=0}^{N-1} \left(\frac{j}{N}\right)^n \leq 0 \\
 N - N \left(1 - \frac{1}{N}\right)^n &\leq \mathbf{E}[M_n] \leq N.
 \end{aligned}$$

Comme $\lim_{n \rightarrow +\infty} \left(1 - \frac{1}{N}\right)^n = 0$, alors d'après le théorème d'encadrement,

$$\lim_{n \rightarrow +\infty} \mathbf{E}[M_n] = N.$$

Remarque. Pour obtenir l'encadrement, il n'y avait pas besoin de la formule sur l'espérance car

* d'une part, $M_n \leq N$, donc $\mathbf{E}[M_n] \leq \mathbf{E}[N] = N$,

* d'autre part, $\mathbf{E}[M_n] = \sum_{k=1}^N k \mathbf{P}(M_n = k) \geq N \mathbf{P}(M_n = N)$,

ce qui donne l'encadrement annoncé. \square

II - Comparaison d'estimateurs

Solution de l'exercice 4.

1. Comme \bar{X}_n et T_n sont des fonctions de X_1, \dots, X_n , alors ce sont des estimateurs.

En utilisant la linéarité de l'espérance,

$$\mathbf{E}[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \lambda = \lambda.$$

Ainsi, \bar{X}_n est un estimateur sans biais de λ .

En utilisant les propriétés de la somme,

$$\begin{aligned} T_n &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - 2\bar{X}_n \sum_{i=1}^n X_i + n\bar{X}_n^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right]. \end{aligned}$$

Ainsi,

$$\begin{aligned} \mathbf{E}[T_n] &= \frac{1}{n-1} \left[\sum_{i=1}^n \mathbf{E}[X_i^2] - n\mathbf{E}[\bar{X}_n^2] \right] \\ &= \frac{n}{n-1} \left[\mathbf{E}[X_1^2] - \mathbf{E}[\bar{X}_n^2] \right] \\ &= \frac{n}{n-1} \left[\mathbf{V}(X_1) + \mathbf{E}[X_1]^2 - \mathbf{V}(\bar{X}_n) - \mathbf{E}[\bar{X}_n]^2 \right]. \end{aligned}$$

Or, comme les X_1, \dots, X_n sont indépendantes,

$$\begin{aligned} \mathbf{V}(\bar{X}_n) &= \mathbf{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathbf{V}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbf{V}(X_i) = \frac{\lambda}{n}. \end{aligned}$$

Finalement,

$$\mathbf{E}[T_n] = \frac{n}{n-1} \left[\lambda + \lambda^2 - \frac{\lambda}{n} - \lambda^2 \right] = \frac{n}{n-1} \frac{n-1}{n} \lambda = \lambda.$$

Ainsi, T_n est un estimateur sans biais de λ .

2. Comme \bar{X}_n est un estimateur sans biais, son risque quadratique est égal à sa variance. Ainsi, d'après les calculs réalisés à la question précédente,

$$R_\lambda(\bar{X}_n) = \mathbf{V}(\bar{X}_n) = \frac{\lambda}{n}.$$

3. Comme $\frac{2\lambda^2}{n-1} > 0$, alors \bar{X}_n (qui est la moyenne empirique) est un estimateur préférable de λ que T_n (qui est l'estimateur sans biais de la variance) car son risque quadratique est plus faible. \square

Solution de l'exercice 5.

1. Comme \bar{X}_n est une fonction de X_1, \dots, X_n , alors \bar{X}_n est un estimateur.

De plus, d'après la linéarité de l'espérance,

$$\mathbf{E}[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i] = \theta.$$

Ainsi, \bar{X}_n est un estimateur sans biais de θ .

Comme X_1, \dots, X_n sont indépendantes, d'après les propriétés de la variance,

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

2. Comme Y_n est une fonction de X_1, \dots, X_n , alors Y_n est un estimateur. De plus, d'après la linéarité de l'espérance,

$$\mathbf{E}[Y_n] = \sum_{i=1}^n \alpha_i \mathbf{E}[X_i] = \theta \sum_{i=1}^n \alpha_i.$$

Ainsi, Y_n est un estimateur sans biais si et seulement si $\sum_{i=1}^n \alpha_i = 1$.

3. En utilisant la bilinéarité de la covariance,

$$\begin{aligned} \text{Cov}(\bar{X}_n, Y_n) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \alpha_j \text{Cov}(X_i, X_j) \\ &= \frac{1}{n} \sum_{i=1}^n \alpha_i \underbrace{\mathbf{V}(X_i)}_{=1}, \text{ d'après l'indépendance} \\ &= \frac{1}{n}. \end{aligned}$$

Ainsi, d'après la positivité de la variance,

$$\begin{aligned} 0 \leq \mathbf{V}(\bar{X}_n - Y_n) &= \mathbf{V}(\bar{X}_n) + \mathbf{V}(Y_n) - 2\text{Cov}(\bar{X}_n, Y_n) \\ &= \frac{1}{n} + \mathbf{V}(Y_n) - 2\frac{1}{n} = \mathbf{V}(Y_n) - \mathbf{V}(\bar{X}_n). \end{aligned}$$

Ainsi, $\mathbf{V}(\bar{X}_n) \leq \mathbf{V}(Y_n)$ avec égalité si et seulement si $\mathbf{V}(\bar{X}_n - Y_n) = 0$, i.e. $\bar{X}_n - Y_n = c$ presque sûrement. Comme $\mathbf{E}[\bar{X}_n] = \mathbf{E}[Y_n] = \theta$, alors $c = 0$ et $\bar{X}_n = Y_n$ presque sûrement.

4. Ainsi, parmi les estimateurs sans biais qui sont des combinaisons linéaires de X_1, \dots, X_n , l'estimateur de θ qui a le plus faible risque quadratique est l'estimateur de la moyenne empirique. \square