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

$$\mathbf{E}[Y_n] = np,$$

$$\mathbf{V}(Y_n) = np(1-p).$$

D'après la formule de Kœnig-Huygens,

$$\mathbf{V}(Y_n) = \mathbf{E}[Y_n^2] - \mathbf{E}[Y_n]^2$$

$$\mathbf{E}[Y_n^2] = \mathbf{V}(Y_n) + \mathbf{E}[Y_n]^2$$

$$= np(1-p) + (np)^2$$

$$= np(1-p+np)$$

$$= np(1+(n-1)p).$$

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

$$\mathbf{E}\left[Y_n^2\right] = np(1 + (n-1)p)$$

$$\mathbf{E}\left[n^2\overline{X}_n^2\right] = np(1 + (n-1)p)$$

$$\mathbf{E}\left[\overline{X}_n^2\right] = \frac{p(1 + (n-1)p)}{n}.$$

Ainsi, \overline{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,

$$\mathbf{E}\left[\overline{X}_{n}^{2}\right] = \frac{p}{n} + \frac{n-1}{n}p^{2}$$

$$\mathbf{E}\left[\overline{X}_{n}^{2}\right] = \frac{\mathbf{E}\left[\overline{X}_{n}\right]}{n} + \frac{n-1}{n}p^{2}$$

$$\mathbf{E}\left[\overline{X}_{n}^{2}\right] - \frac{1}{n}\mathbf{E}\left[\overline{X}_{n}\right] = \frac{n-1}{n}p^{2}$$

$$\frac{n}{n-1}\mathbf{E}\left[\overline{X}_{n}^{2} - \frac{1}{n}\overline{X}_{n}\right] = p^{2}$$

$$\mathbf{E}\left[\frac{n}{n-1}\overline{X}_{n}^{2} - \frac{1}{n-1}\overline{X}_{n}\right] = p^{2}.$$

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

2º méthode. En utilisant la variance empirique, $\overline{X}_n - \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ 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

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

$$= \frac{1}{n-1} \sum_{i=1}^{n} X_i - \frac{2}{n-1} \overline{X}_n \sum_{i=1}^{n} X_i + \frac{n}{n-1} \overline{X}_n^2$$

$$= \frac{n}{n-1} \overline{X}_n - \frac{2n}{n-1} \overline{X}_n^2 + \frac{n}{n-1} \overline{X}_n^2$$

$$= \frac{n}{n-1} \overline{X}_n - \frac{n}{n-1} \overline{X}_n^2$$

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

$$= \frac{n}{n-1} \overline{X}_n^2 - \frac{1}{n-1} \overline{X}_n.$$

Solution de l'exercice 2.

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

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

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

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

$$\mathbf{V}\left(\sum_{i=1}^{n} X_{i}\right) = \sum_{i=1}^{n} \mathbf{V}\left(X_{i}\right) = \sum_{i=1}^{n} \sigma^{2} = n\sigma^{2}$$
$$\mathbf{V}\left(\overline{X}_{n}\right) = \mathbf{V}\left(\frac{1}{n}\overline{X}_{n}\right) = \frac{1}{n^{2}}\mathbf{V}\left(\overline{X}_{n}\right) = \frac{\sigma^{2}}{n}.$$

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

$$ns_n = \sum_{i=1}^n (X_i - \overline{X}_n)^2$$

$$= \sum_{i=1}^n \left(X_i^2 - 2X_i \overline{X}_n + \overline{X}_n^2 \right)$$

$$= \sum_{i=1}^n X_i^2 - 2\overline{X}_n \sum_{i=1}^n X_i + \sum_{i=1}^n \overline{X}_n^2$$

$$= \sum_{i=1}^n X_i^2 - 2n\overline{X}_n^2 + n\overline{X}_n^2$$

$$s_n = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}_n^2.$$

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

$$\mathbf{E}\left[s_{n}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}\left[X_{i}^{2}\right] - \mathbf{E}\left[\overline{X}_{n}^{2}\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{V}\left(X_{i}\right) + \mathbf{E}\left[X_{i}\right]^{2}\right) - \left(\mathbf{V}\left(\overline{X}_{n}\right) + \mathbf{E}\left[\overline{X}_{n}\right]^{2}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\sigma^{2} + m^{2}) - \left(\frac{\sigma^{2}}{n} + m^{2}\right)$$

$$= \sigma^{2} + m^{2} - \left(\frac{\sigma^{2}}{n} + m^{2}\right)$$

$$= \frac{n-1}{n} \sigma^{2}.$$

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, \ldots, 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

D 2

maximum:

$$\mathbf{P}\left([M_n \leqslant i]\right) = \mathbf{P}\left(\left[\max_{k=1} \left\{X_1, \dots, X_n\right\} \leqslant i\right]\right)$$

$$= \prod_{k=1}^n \mathbf{P}\left([X_k \leqslant i]\right), \text{ en utilisant l'indépendance}$$

$$= \prod_{k=1}^n \mathbf{P}\left([X_1 \leqslant i]\right), \text{ car les tirages sont avec remises}$$

$$= \prod_{k=1}^n \left[X_1 \leqslant i\right]^n$$

$$= \left(\sum_{\ell=1}^i \mathbf{P}\left([X_1 = \ell]\right)\right)^n, \text{ car } X_1 \text{ prend les valeurs } 1, \dots, N$$

$$= \left(\frac{i}{n}\right)^n, \text{ car les boules sont indistinguables}$$

2. En utilisant la formule précédente,

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

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

 ${\bf 3.}\,$ On va effectuer un changement dans l'ordre des sommations :

$$\sum_{i=1}^{N} \mathbf{P}(Y \geqslant i) = \sum_{i=1}^{N} \sum_{j=i}^{N} \mathbf{P}(Y = j)$$

$$= \sum_{1 \leq i \leq j \leq N} \mathbf{P}(Y = j)$$

$$= \sum_{j=1}^{N} \sum_{i=1}^{j} \mathbf{P}(Y = j)$$

$$= \sum_{j=1}^{N} j \mathbf{P}(Y = j) = \mathbf{E}[Y].$$

4. En utilisant les questions précédentes, comme M_n est à valeurs dans $[\![1,N]\!],$

$$\mathbf{E}[M_n] = \sum_{i=1}^{N} \mathbf{P}([Y \geqslant i])$$

$$= \sum_{i=1}^{N} (1 - \mathbf{P}([Y \leqslant i - 1]))$$

$$= \sum_{i=1}^{N} \left(1 - \left(\frac{i-1}{N}\right)^n\right)$$

$$= N - \sum_{i=1}^{N} \left(\frac{i-1}{N}\right)^n$$

$$= N - \sum_{i=0}^{N-1} \left(\frac{i}{N}\right)^n.$$

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

$$\frac{0}{N} \leqslant \frac{j}{N} \leqslant \frac{N-1}{N}$$

$$0 \leqslant \left(\frac{j}{N}\right)^n \leqslant \left(\frac{N-1}{N}\right)^n$$

$$-\left(\frac{N-1}{N}\right)^n \leqslant \left(\frac{j}{N}\right)^n \leqslant 0$$

$$-\sum_{j=0}^{N-1} \left(1 - \frac{1}{N}\right)^n \leqslant \sum_{j=0}^{N-1} \left(\frac{j}{N}\right)^n \leqslant 0$$

$$N - N\left(1 - \frac{1}{N}\right)^n \leqslant \mathbf{E}\left[M_n\right] \leqslant N.$$

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

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

A. Camanes

II - Comparaison d'estimateurs

Solution de l'exercice 4.

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

En utilisant la linéarité de l'espérance,

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

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

En utilisant les propriétés de la somme,

$$T_{n} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i} - \overline{X}_{n})^{2}$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_{i}^{2} - 2X_{i}\overline{X}_{n} + \overline{X}_{n}^{2})$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_{i}^{2} - 2\overline{X}_{n} \sum_{i=1}^{n} X_{i} + n\overline{X}_{n}^{2} \right]$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} X_{i}^{2} - n\overline{X}_{n}^{2} \right].$$

Ainsi,

$$\mathbf{E}\left[T_{n}\right] = \frac{1}{n-1} \left[\sum_{i=1}^{n} \mathbf{E}\left[X_{i}^{2}\right] - n\mathbf{E}\left[\overline{X}_{n}^{2}\right] \right]$$

$$= \frac{n}{n-1} \left[\mathbf{E}\left[X_{1}^{2}\right] - \mathbf{E}\left[\overline{X}_{n}^{2}\right] \right]$$

$$= \frac{n}{n-1} \left[\mathbf{V}\left(X_{1}\right) + \mathbf{E}\left[X_{1}\right]^{2} - \mathbf{V}\left(\overline{X}_{n}\right) - \mathbf{E}\left[\overline{X}_{n}\right]^{2} \right].$$

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

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

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 \overline{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,

$$R_{\lambda}\left(\overline{X}_{n}\right) = \mathbf{V}\left(\overline{X}_{n}\right) = \frac{\lambda}{n}.$$

3. Comme $\frac{2\lambda^2}{n-1} > 0$, alors \overline{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.

Solution de l'exercice 5.

1. Commme \overline{X}_n est une fonction de X_1, \ldots, X_n , alors \overline{X}_n est un estimateur.

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

$$\mathbf{E}\left[\overline{X}_{n}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbf{E}\left[X_{i}\right] = \theta.$$

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

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

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

2. Comme Y_n est une fonction de X_1, \ldots, 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,

$$\operatorname{Cov}\left(\overline{X}_{n}, Y_{n}\right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{j} \operatorname{Cov}\left(X_{i}, X_{j}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \alpha_{i} \underbrace{\mathbf{V}\left(X_{i}\right)}_{=1}, \text{ d'après l'indépendance}$$

$$= \frac{1}{n}.$$

Ainsi, d'après la positivité de la variance,

$$0 \leq \mathbf{V}\left(\overline{X}_{n} - Y_{n}\right) = \mathbf{V}\left(\overline{X}_{n}\right) + \mathbf{V}\left(Y_{n}\right) - 2\operatorname{Cov}\left(\overline{X}_{n}, Y_{n}\right)$$
$$= \frac{1}{n} + \mathbf{V}\left(Y_{n}\right) - 2\frac{1}{n} = \mathbf{V}\left(Y_{n}\right) - \mathbf{V}\left(\overline{X}_{n}\right).$$

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

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