Ex 16 Soit une variable aléatoire X de loi de Poisson de paramètre λ ($\lambda > 0$). Le but de cet exercice est de trouver une estimation de $\theta = e^{-\lambda}$.
$(X_1, X_2, \ldots, X_n)$ un échantillon de X.
et $Y_1, Y_2, \ldots, Y_n$ des variables aléatoires définies par $Y_i = 1$ si $X_i = 0$ et $Y_i = 0$ sinon.
$\forall i \in [\![1, n]\!]$

$$S_n = \sum_{i=1}^{n} X_i \quad , \quad \overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \quad , \quad T_n = \left(\frac{n-1}{n}\right)^{S_n}.$$

1) Montrer que $\overline{Y}_n$ est un estimateur sans biais et convergent de $\theta = e^{-\lambda}$.

2) Montrer que T_n est un estimateur sans biais et convergent de θ

3) a) Etudier le sens de variation de f définie sur $\mathbb{R}_+$ par $f(t) = n e^{t/n} - e^{t} - n + 1$

et en déduire son signe

b) En déduire que T_n est meilleur estimateur que $\overline{Y}_n$

Ex 16

1) Y_i suit la loi de Bernoulli, le paramètre de Y_i est $P(Y_i = 1) = P(X_i = 0) = e^{-1} = \theta$

Donc $Y_i \hookrightarrow \mathcal{B}(\theta)$

$$E(\bar{Y}_n) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{n\theta}{n} = \theta$$

$\bar{Y}_n$ est sans biais

$$V(\bar{Y}_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(Y_i) = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n} \xrightarrow[n\to+\infty]{} 0$$

$$V(\bar{Y}_n) \xrightarrow[n\to+\infty]{} 0$$

En appliquant Tchebychev : $\forall \varepsilon > 0$

$$P(|\bar{Y}_n - E(\bar{Y}_n)| \geq \varepsilon) \leq \frac{V(\bar{Y}_n)}{\varepsilon^2} = \frac{\theta(1-\theta)}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0$$

Donc $\bar{Y}_n \xrightarrow[n\to+\infty]{P} \theta$

2) $T_n = \left(\frac{n-1}{n}\right)^{S_n}$

$$E(T_n) = E\left(\left(\frac{n-1}{n}\right)^{S_n}\right) = \sum_{k=0}^{+\infty}\left(\frac{n-1}{n}\right)^{k} P(S_n = k)$$

$$\left(\text{car } E(\varphi(X)) = \sum_{k} \varphi(k) P(X = k)\right)$$

$$E(T_n) = \sum_{k=0}^{+\infty} \left(\frac{n-1}{n}\right)^k e^{-n\lambda} \frac{(n\lambda)^k}{k!}$$

(Car $S_n = \sum_{i=1}^{n} X_i$ somme indépendantes de Poisson $\mathcal{P}(\lambda)$

donc $S_n \longrightarrow \mathcal{P}(n\lambda)$)

$$E(T_n) = e^{-n\lambda} \sum_{k=0}^{+\infty} \frac{((n-1)\lambda)^k}{k!} = e^{-n\lambda} \cdot e^{(n-1)\lambda} = e^{-\lambda} = \theta$$

(Rappel $\sum_{0}^{+\infty} \frac{x^k}{k!} = e^x$)

T_n est sans biais.

$$E(T_n^2) = E\left(\left(\frac{n-1}{n}\right)^{2S_n}\right) = \sum_{0}^{+\infty} \left(\frac{n-1}{n}\right)^{2k} P(S_n = k)$$

$$E(T_n^2) = \sum_{0}^{+\infty} \left(\frac{n-1}{n}\right)^{2k} e^{-n\lambda} \frac{(n\lambda)^k}{k!}$$

$$E(T_n^2) = e^{-n\lambda} \sum_{k=0}^{+\infty} \frac{\left(\frac{(n-1)^2\lambda}{n}\right)^k}{k!}$$

$$= e^{-n\lambda} \cdot e^{(n-1)^2 \frac{\lambda}{n}} = e^{-n\lambda} \cdot e^{(n^2-2n+1)\frac{\lambda}{n}}$$

$$= e^{-2\lambda + \frac{\lambda}{n}} = \theta^2 e^{\frac{\lambda}{n}}$$

(20)

Donc $\boxed{E(T_n^2) = \theta^2 e^{\lambda/n}}$

$V(T_n) = E(T_n^2) - E^2(T_n) = \theta^2 e^{\lambda/n} - \theta^2$
$= \theta^2 (e^{\lambda/n} - 1)$

$\lim_{n \to +\infty} V(T_n) = 0$

En utilisant Tchebychev : $\forall \varepsilon > 0$

$P(|T_n - \theta| \geq \varepsilon) < \frac{V(T_n)}{\varepsilon^2} \xrightarrow[n \to +\infty]{} 0$

$\Rightarrow T_n \xrightarrow[n \to +\infty]{P} \theta$

3) a) $f(t) = n e^{t/n} - e^{t} - n + 1$
$\forall t \in \mathbb{R}_+$

$f'(t) = e^{t/n} - e^{t}$

$\forall n \geq 1$ $\frac{t}{n} \leq t \Rightarrow e^{t/n} \leq e^{t} \Rightarrow f'(t) \leq 0$
$\forall t > 0$ $\forall t \in \mathbb{R}_+$

$f(t)$ est décroissante

et comme $f(0) = 0 \Rightarrow \forall t \geq 0$
$f(t) \leq f(0) = 0$

$\Rightarrow \boxed{f(t) \leq 0 \quad \forall t \geq 0}$

(21)

5) $E(T_n) = \theta$ \} les deux estimateurs de θ
$E(\overline{Y}_n) = \theta$ \} sont sans biais

Comparons leurs variances.

$$V(\overline{Y}_n) = \frac{\theta(1-\theta)}{n}, \quad V(T_n) = \theta^2\left(e^{\frac{\lambda}{n}} - 1\right)$$

$$V(T_n) - V(\overline{Y}_n) = \theta^2\left(e^{\lambda/n} - 1\right) - \frac{\theta(1-\theta)}{n}$$

$$= \frac{\theta^2}{n}\left(n e^{\lambda/n} - n - \frac{1}{\theta} + 1\right)$$

$$V(T_n) - V(\overline{Y}_n) = \frac{\theta^2}{n}\left(n e^{\lambda/n} - n - e^{\lambda} + 1\right)$$

$$= \frac{\theta^2}{n} f(\lambda)$$

or f est négative $\Rightarrow V(T_n) - V(\overline{Y}_n) \leq 0$

$$\Rightarrow \boxed{V(T_n) \leq V(\overline{Y}_n)}$$

Conclusion : T_n est un meilleur estimateur que $\overline{Y}_n$.

Ex 17 $X \hookrightarrow \mathcal{B}(N,p)$ $\theta = p$ inconnu.

1) $L(x_1, x_2, \dots, x_n, p) = \dfrac{(N!)^n}{\prod_{i=1}^{n} x_i!(N-x_i)!} \cdot p^{\sum_{i=1}^{n} x_i} (1-p)^{nN-\sum_{i=1}^{n} x_i}$

(d'après l'exercice 14)

2) L'équation de la vraisemblance :

$$\frac{\partial}{\partial p} \ln L = 0$$

$$\ln L(x_1, \dots, x_n, p) = \ln\left(\frac{(N!)^n}{\prod_{i=1}^{n} x_i!(N-x_i)!}\right) + \sum_{i=1}^{n} x_i \ln p + \left(nN - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

$$\frac{\partial \ln L}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i + \left(nN - \sum_{i=1}^{n} x_i\right)\frac{(-1)}{1-p} = 0$$

$$\Leftrightarrow (1-p)\sum_{i=1}^{n} x_i - p\left(nN - \sum_{i=1}^{n} x_i\right) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i - pnN = 0$$

$$\Leftrightarrow \boxed{\hat{p} = \frac{1}{nN}\sum_{i=1}^{n} x_i}$$ estimation ponctuelle de p.

~~L'estimateur~~ L'estimateur du maximum de vraisemblance est $\boxed{T_n = \frac{1}{nN}\sum_{i=1}^{n} X_i}$ (23)

3) T_n sans biais ?

$$E(T_n) = \frac{1}{nN}\sum_{i=1}^{n} E(X_i) = \frac{1}{nN}\sum_{i=1}^{n} Np = \frac{nNp}{nN} = p$$

$\boxed{E(T_n) = p}$ T_n est sans biais

4) Convergence ?

$$V(T_n) = \frac{1}{n^2N^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2N^2}\sum_{i=1}^{n} Np(1-p)$$

$$V(T_n) = \frac{nNp(1-p)}{n^2N^2} = \frac{p(1-p)}{nN} \xrightarrow[n \to +\infty]{} 0$$

D'après Tchebychev :

$$\forall \varepsilon > 0 \quad P(|T_n - E(T_n)| \geq \varepsilon) \leq \frac{V(T_n)}{\varepsilon^2}$$

$$\Rightarrow P(|T_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{nN\varepsilon^2} \xrightarrow[n \to +\infty]{} 0$$

Donc $\boxed{T_n \xrightarrow[n \to +\infty]{P} p}$

T_n converge en probabilité vers p. (26)

5) Efficacité

$$I_n(p) = -E\left(\frac{\partial^2 \ln L}{\partial p^2}\right)$$

Information de Fisher

$$\frac{\partial \ln L}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \left(nN - \sum_{i=1}^{n} x_i\right)\frac{1}{1-p}$$

$$\frac{\partial^2 \ln L}{\partial p^2} = -\frac{1}{p^2}\sum_{i=1}^{n} x_i + \left(nN - \sum_{i=1}^{n} x_i\right)\frac{(-1)}{(1-p)^2}$$

$$E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) = -\frac{1}{p^2}\sum_{i=1}^{n} E(x_i) + \frac{(-1)}{(1-p)^2}\left(nN - \sum_{i=1}^{n} E(x_i)\right)$$

$$= -\frac{1}{p^2} nNp + \frac{1}{(1-p)^2}\left(-nN + nNp\right)$$

$$E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) = -\frac{nN}{p} + \frac{(-nN)}{1-p} = \frac{-nN(1-p) - nNp}{p(1-p)}$$

$$E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) = \frac{-nN}{p(1-p)}$$

$$I_n(p) = -E\left(\frac{\partial^2 \ln L}{\partial p^2}\right) = \frac{nN}{p(1-p)}$$

Donc $\boxed{I_n(p) = \frac{nN}{p(1-p)}}$ Information de Fisher

or $V(T_n) = \frac{p(1-p)}{nN} \Rightarrow \boxed{V(T_n) = \frac{1}{I_n(p)}}$

Conclusion T_n est efficace

Ex 18 $X \nearrow \mathcal{P}(\theta)$ Poisson de paramètre θ

θ : inconnu

1) La vraisemblance est :

$$L(x_1, x_2, \ldots, x_n, \theta) = \prod_{i=1}^{n} P(X_i = x_i)$$

$$L(x_1, \ldots, x_n, \theta) = \frac{e^{-n\theta} \, \theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \quad \text{(Exercice 14)}$$

2) Méthode du maximum de vraisemblance :

$$\frac{\partial \ln L}{\partial \theta} = 0 \quad \text{(Eq. de la vraisemblance)}$$

Ex 18 (suite)

$$\ln L(x_1,\dots,x_n,\theta) = -n\theta + \sum_{i=1}^{n} x_i \ln\theta - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

$$\frac{\partial \ln L}{\partial \theta} = 0 \Leftrightarrow -n + \frac{1}{\theta}\sum_{i=1}^{n} x_i = 0$$

$$\Leftrightarrow \boxed{\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$ estimation ponctuelle de θ.

L'estimateur de θ est $T_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

3) Tn sans biais ?

$$E(T_n) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\theta = \frac{n\theta}{n} = \theta$$

$\boxed{E(T_n) = \theta}$ T_n est sans biais

Convergence ?

$$V(T_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n\theta}{n^2} = \frac{\theta}{n}$$

$$V(T_n) = \frac{\theta}{n} \xrightarrow[n \to +\infty]{} 0$$

Donc en utilisant Tchebychev :

$$\forall \varepsilon > 0, \ P\left(|T_n - E(T_n)| \geq \varepsilon\right) \leq \frac{V(T_n)}{\varepsilon^2}$$

$$\Rightarrow P\left(|T_n - \theta| \geq \varepsilon\right) \leq \frac{\theta}{n\varepsilon^2} \xrightarrow[n \to +\infty]{} 0$$

Donc $T_n \xrightarrow[n \to +\infty]{P} \theta$

4) Efficacité

On calcule l'information de Fisher :

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$$

$$\frac{\partial \ln L}{\partial \theta} = -n + \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{1}{\theta^2}\sum_{i=1}^{n} x_i$$

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = \frac{1}{\theta^2}\sum_{i=1}^{n} E(x_i) = \frac{n\theta}{\theta^2}$$

$$\boxed{I_n(\theta) = \frac{n}{\theta}}$$

Donc $V(T_n) = \frac{\theta}{n} = \frac{1}{I_n(\theta)} \Rightarrow T_n$ est efficace

Ex 19

1) f étant une densité : $\int_{\mathbb{R}} f(x)dx = 1$

$$A \int_1^{+\infty} \frac{1}{x^{1+1/\theta}} dx = 1$$

$$A \left[\frac{-\theta}{x^{1/\theta}} \right]_1^{+\infty} = 1 \Rightarrow A\theta = 1 \Rightarrow \boxed{A = \frac{1}{\theta}}$$

2) $L(x_1, x_2, .., x_n, \theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{1}{\theta} \cdot \frac{1}{x_i^{1+1/\theta}}$

$$\boxed{L(x_1, x_2, .., x_n, \theta) = \frac{1}{\theta^n} \frac{1}{\prod_{i=1}^{n} x_i^{1+1/\theta}}}$$

3) $\ln L(x_1, .., x_n, \theta) = -n \ln \theta - \left(1 + \frac{1}{\theta}\right) \sum_{i=1}^{n} \ln x_i$

Eq. de la vraisemblance :

$$\frac{\partial \ln L}{\partial \theta} = 0$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} \ln x_i = 0$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \ln x_i}$$ estimateur ponctuelle

(29)

L'estimateur du maximum de vraisemblance :

$$\boxed{T_n = \frac{1}{n} \sum_{i=1}^{n} \ln X_i}$$

4) $E(T_n) = \frac{1}{n} \sum_{i=1}^{n} E(\ln X_i) = \frac{n E(\ln X)}{n} = E(\ln X)$

or $E(\ln X) = \int_1^{+\infty} \ln x \, f(x) dx = \int_1^{+\infty} \frac{\ln x}{\theta x^{1+1/\theta}} dx$

on intègre par partie :

$\begin{cases} v = \ln x & v' = \frac{1}{x} \\ u' = \frac{1}{\theta} x^{-1-1/\theta} & u = -x^{-1/\theta} \end{cases}$

$E(\ln x) = \left[-x^{-1/\theta} \ln x \right]_1^{+\infty} + \int_1^{+\infty} x^{-1-1/\theta} dx$

$\left[-x^{-1/\theta} \ln x \right] \xrightarrow{x \to +\infty} 0$

$E(\ln x) = \left[-\theta x^{-1/\theta} \right]_1^{+\infty} = \theta$

Donc $\boxed{E(T_n) = \theta}$ Sans biais

convergence ? $V(T_n) = \frac{1}{n^2} \sum_{i=1}^{n} V(\ln X_i) = \frac{n V(\ln X)}{n^2}$

$= \frac{V(\ln x)}{n}$

(3c)

$$V(T_n) = \frac{V(\ln X)}{n}$$

$$E(\ln^2 X_i) = \int_1^{+\infty} \frac{\ln^2 x}{\theta x^{1+1/\theta}}\,dx = \underbrace{\left[-x^{-1/\theta}\ln^2 x\right]_1^{+\infty}}_{x\to+\infty\ \ 0} + \int_1^{+\infty} \frac{2\ln x}{x^{1+1/\theta}}\,dx$$

$$\begin{cases} v = \ln^2 x & v' = 2(\ln x)\frac{1}{x} \\ u' = \frac{1}{\theta} x^{-1-\frac{1}{\theta}}, & u = -x^{-\frac{1}{\theta}} \end{cases}$$

$$E(\ln^2 x) = 2\theta \int_1^{+\infty} \frac{\ln x\,dx}{\theta x^{1+1/\theta}} = 2\theta E(\ln x) = 2\theta^2$$

$$V(T_n) = \frac{E(\ln^2 x) - E^2(\ln x)}{n}$$

$$V(T_n) = \frac{1}{n}(2\theta^2 - \theta^2) = \frac{\theta^2}{n}$$

$$V(T_n) = \frac{\theta^2}{n} \xrightarrow[n\to+\infty]{} 0$$

D'après Tchebychev $T_n \xrightarrow[n\to+\infty]{P} \theta$

5) <u>Efficacité</u>

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right)$$

(51)

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = +\frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n} \ln x_i$$

$$I_n(\theta) = -E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) = -\frac{n}{\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n} E(\ln x_i)$$

$$I_n(\theta) = -\frac{n}{\theta^2} + \frac{2}{\theta^3} n E(\ln x)$$

$$I_n(\theta) = \frac{-n}{\theta^2} + \frac{2}{\theta^3} n\theta = \boxed{\frac{n}{\theta^2}}$$

or $V(T_n) = \frac{\theta^2}{n} = \frac{1}{I_n(\theta)}$

Donc T_n est efficace