

Exercice 1

(1)

1) La moyenne d'une loi binomiale $B(k, p)$ est

$$\begin{aligned} E[X] &= \sum_{i=0}^k i P[X=i] = \sum_{i=0}^k i \frac{k!}{i!(k-i)!} p^i (1-p)^{k-i} \\ &= \sum_{i=1}^k \frac{k!}{(i-1)!(k-i)!} p^i (1-p)^{k-i} \\ &\stackrel{j=i-1}{=} \sum_{j=0}^{k-1} \frac{k!}{j!(k-j-1)!} p^{j+1} (1-p)^{k-j-1} \\ &= kp \sum_{j=0}^{k-1} \frac{(k-1)!}{j!(k-1-j)!} p^j (1-p)^{k-1-j} \\ &= kp (p+1-p)^{k-1} \\ &= \boxed{kp} \quad \text{ou} \quad \boxed{p = \frac{1}{k} E[X]} \end{aligned}$$

donc

$$\hat{p}_{\text{MOM}} = \frac{1}{k} \frac{1}{n} \sum_{i=1}^n x_i$$

2) La vraisemblance de (x_1, \dots, x_n) s'écrit

$$L(x_1, \dots, x_n) = P[X_1=x_1, \dots, X_n=x_n] = \prod_{i=1}^n C_k^{x_i} p^{x_i} (1-p)^{k-x_i}$$

Le logarithme de la vraisemblance s'en déduit

$$\ln L(x_1, \dots, x_n) = \sum_{i=1}^n \ln(C_k^{x_i}) + x_i \ln p + (k-x_i) \ln(1-p)$$

On a donc

$$\frac{\partial \ln L(x_1, \dots, x_n)}{\partial p} = \frac{1}{p} \sum_{i=1}^n x_i + \sum_{i=1}^n (k-x_i) \frac{(-1)}{1-p}$$

En posant $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, on obtient les variations suivantes

$$\frac{\partial \ln L(x_1, \dots, x_n)}{\partial p} \geq 0 \Leftrightarrow \frac{n}{p} \bar{x} \geq \frac{1}{1-p} \sum_{i=1}^n (k-x_i) = \frac{n(k - n\bar{x})}{1-p}$$

$$\Leftrightarrow n\bar{x}(1-p) \geq np(k - \bar{x})$$

$$\Leftrightarrow n\bar{x} \geq npk$$

$$\Leftrightarrow \boxed{p \leq \frac{1}{k} \bar{x}}$$

On a donc le tableau de variations suivant

(2)

p	0	$\frac{1}{k}\bar{x}$	1
$\frac{\partial \ln L}{\partial p}$	+	0	-
$\ln L$			

$\frac{1}{k}\bar{x}$ est donc le maximum global unique de la log-vraisemblance, d'où

$$\hat{p}_{MV} = \frac{1}{k} \frac{1}{n} \sum_{i=1}^n x_i$$

Pour une loi binomiale, les estimateurs du maximum de vraisemblance et des moments (utilisant la moyenne) sont identiques.

on a $E[\hat{p}_{MV}] = \frac{1}{kn} \sum_{i=1}^n \underbrace{E[x_i]}_{kp} = p$

$\text{Var } \hat{p}_{MV} = \frac{1}{k^2 n^2} \sum_{i=1}^n \underbrace{\text{Var } x_i}_{kp(1-p)}$ car les x_i sont indépendantes

d'où $\text{Var } \hat{p}_{MV} = \frac{p(1-p)}{nk}$

L'estimateur \hat{p}_{MV} est donc non biaisé. Puisque $\text{Var } \hat{p}_{MV} \xrightarrow[n \rightarrow +\infty]{} 0$, l'estimateur \hat{p}_{MV} est convergent.

3) D'après la question précédente

$$\frac{\partial^2 \ln L(x_1, \dots, x_n)}{\partial p^2} = -\frac{1}{p^2} \sum_{i=1}^n x_i + \sum_{i=1}^n (k - x_i) \left(\frac{-1}{(1-p)^2} \right)$$

$$\begin{aligned} \text{donc } E\left[\frac{\partial^2 \ln L(x_1, \dots, x_n)}{\partial p^2}\right] &= -\frac{1}{p^2} \sum_{i=1}^n \underbrace{E[x_i]}_{kp} - \frac{1}{(1-p)^2} \sum_{i=1}^n \left(k - \underbrace{E[x_i]}_{kp} \right) \\ &= -\frac{nk}{p} - \frac{nk(1-p)}{(1-p)^2} \\ &= \frac{-nk(1-p) - nk p}{p(1-p)} = \frac{-nk}{p(1-p)} \end{aligned}$$

La borne de Cramer-Rao pour un estimateur non biaisé de p est donc

$$BCR(p) = \frac{-1}{E\left[\frac{\partial^2 \ln L(x_1, \dots, x_n)}{\partial p^2}\right]} = \boxed{\frac{p(1-p)}{nk}}$$

L'estimateur \hat{p}_{MV} vérifie $Var(\hat{p}_{MV}) = BCR(p)$ et il est non biaisé donc \hat{p}_{MV} est l'estimateur efficace de p

4) • La loi a posteriori de p est telle que

$$\begin{aligned} f(p|x_1, \dots, x_n) &\propto f(x_1, \dots, x_n | p) f(p) \\ &\propto p^{\sum_{i=1}^n x_i} (1-p)^{nk - \sum_{i=1}^n x_i} p^{\alpha-1} (1-p)^{\beta-1} \end{aligned}$$

c'est-à-dire

$$f(p|x_1, \dots, x_n) \propto p^{n\bar{x} + \alpha - 1} (1-p)^{nk - n\bar{x} + \beta - 1}$$

qui est une loi Beta

$$\boxed{Be(n\bar{x} + \alpha, nk - n\bar{x} + \beta)}$$

• La moyenne de la loi a posteriori de p est

$$E[p|x_1, \dots, x_n] = \frac{n\bar{x} + \alpha}{\alpha + \beta + nk}$$

L'estimateur MMSE de p est

$$\boxed{\hat{p}_{MMSE} = \frac{\sum_{i=1}^n x_i + \alpha}{\alpha + \beta + nk}}$$

quand n est "petit", \hat{p}_{MMSE} est proche de $\frac{\alpha}{\alpha + \beta}$ qui est la moyenne de la loi a priori

quand n est "grand", \hat{p}_{MMSE} est proche de $\frac{\sum_{i=1}^n x_i}{nk}$ qui est l'estimateur du maximum de vraisemblance de p

Dans le cas général, \hat{p}_{MMSE} est une combinaison linéaire des deux estimateurs énoncés ci-dessus :

$$\hat{p}_{MMSE} = \frac{1}{1 + \frac{\alpha + \beta}{nk}} \left(\frac{\sum x_i}{nk} \right) + \frac{1}{1 + \frac{nk}{\alpha + \beta}} \left(\frac{\alpha}{\alpha + \beta} \right)$$

- L'estimateur MAP s'obtient en maximisant le logarithme de la loi a posteriori de p , i.e en maximisant (4)

$$\pi(p) = (n\bar{x} + \alpha - 1) \ln p + (n\kappa - n\bar{x} + \beta - 1) \ln(1-p)$$

ona $\pi'(p) = \frac{n\bar{x} + \alpha - 1}{p} + \frac{n\kappa - n\bar{x} + \beta - 1}{p-1}$

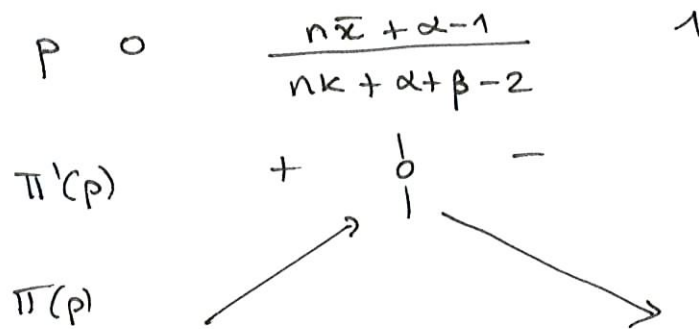
donc $\pi'(p) \geq 0 \Leftrightarrow \frac{n\bar{x} + \alpha - 1}{p} \geq \frac{n\kappa - n\bar{x} + \beta - 1}{1-p}$

$$\Leftrightarrow (1-p)(n\bar{x} + \alpha - 1) \geq (n\kappa - n\bar{x} + \beta - 1)p$$

$$\Leftrightarrow n\bar{x} + \alpha - 1 \geq p(n\kappa - n\bar{x} + \beta - 1 + n\bar{x} + \alpha - 1)$$

$$\Leftrightarrow p \leq \frac{n\bar{x} + \alpha - 1}{n\kappa + \alpha + \beta - 2}$$

d'où les variations suivantes



L'estimateur MAP de p est donc

$$\hat{p}_{\text{MAP}} = \frac{\sum_{i=1}^n x_i + \alpha - 1}{n\kappa + \alpha + \beta - 2}$$

• quand n est "grand", \hat{p}_{MAP} se comporte comme $\frac{\sum_{i=1}^n x_i}{n\kappa}$ qui est l'estimateur du maximum de vraisemblance de p .

• quand n est "petit", \hat{p}_{MAP} se comporte comme $\frac{\alpha - 1}{\alpha + \beta - 2}$ qui est

le mode de la loi a priori (la valeur de p qui maximise la loi a priori de p)

Dans le cas général, \hat{p}_{MAP} est une combinaison linéaire des 2 estimateurs ci-dessus

$$\hat{p}_{\text{MAP}} = \frac{1}{1 + \frac{\alpha + \beta - 2}{n\kappa}} \left(\frac{\sum_{i=1}^n x_i}{n\kappa} \right) + \frac{1}{1 + \frac{n\kappa}{\alpha + \beta - 2}} \left(\frac{\alpha - 1}{\alpha + \beta - 2} \right)$$

1) Le test de Neyman-Pearson s'écrit

$$\text{Rejet de } H_0 \text{ si } \frac{L(x_1, \dots, x_n; p_1)}{L(x_1, \dots, x_n; p_0)} > S_\alpha$$

$$\text{si } \frac{p_1^{\sum x_i} (1-p_1)^{nk - \sum x_i}}{p_0^{\sum x_i} (1-p_0)^{nk - \sum x_i}} > S_\alpha$$

$$\text{si } (\sum x_i) \left[\underbrace{\ln p_1 - \ln(1-p_1) - \ln p_0 + \ln(1-p_0)}_{\ln \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right]} \right] > k_\alpha$$

• Si $p_1 > p_0$, alors $\frac{p_1}{p_0} > 1$ et $\frac{1-p_0}{1-p_1} > 1$ donc $\ln \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right] > 0$

$$\text{d'où } \boxed{\text{rejet de } H_0 \text{ si } \sum_{i=1}^n x_i > V_\alpha}$$

$$\text{donc } \boxed{T_n = \sum_{i=1}^n x_i}$$

• Si $p_1 < p_0$, alors $\boxed{\text{on rejette } H_0 \text{ si } \sum_{i=1}^n x_i < V_\alpha}$

2) On sait que la loi binomiale correspond au nombre de succès sur k expériences de type succès / échec. Si Y_{ij} est la variable aléatoire définie par $Y_{ij} = \begin{cases} 1 & \text{si il y a succès pour la } j^{\text{ème}} \text{ expérience} \\ 0 & \text{si il y a échec pour la } j^{\text{ème}} \text{ expérience} \end{cases}$

$$\text{alors } \boxed{X_i = \sum_{j=1}^k Y_{ij}}$$

On en déduit $T_n = \sum_{i=1}^n X_i = \sum_{i=1}^n \sum_{j=1}^k Y_{ij}$ qui est la somme de

nk variables succès / échec donc T_n suit une loi binomiale

$$\underline{B(nk, p)}$$

3) Le théorème de la limite centrale s'écrit

(6)

$$\frac{\sum_{i=1}^n X_i - nkp}{nkp(1-p)} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0,1)$$

Pour n "grand", on peut donc approcher la loi de $T_n = \sum_{i=1}^n X_i$ par une loi normale

$$\boxed{\mathcal{N}(nkp, nkp(1-p))}$$

Sous H_0 , $p = p_0$ et sous H_1 , $p = p_1$ donc

$$T_n | H_0 \sim \mathcal{N}(nkp_0, nkp_0(1-p_0))$$

$$T_n | H_1 \sim \mathcal{N}(nkp_1, nkp_1(1-p_1))$$

4) on a

$$\alpha = P[\text{Rejeter } H_0 | H_0 \text{ vraie}] = P[T_n > \gamma_\alpha | H_0 \text{ vraie}]$$

$$\text{d'où } \alpha = P\left[\frac{T_n - nkp_0}{\sqrt{nkp_0(1-p_0)}} > \frac{\gamma_\alpha - nkp_0}{\sqrt{nkp_0(1-p_0)}} \mid U \sim \mathcal{N}(0,1) \right]$$

$$\text{d'où } \alpha = 1 - \Phi\left[\frac{\gamma_\alpha - nkp_0}{\sqrt{nkp_0(1-p_0)}} \right]$$

$$\text{d'où } \boxed{\gamma_\alpha = \sqrt{nkp_0(1-p_0)} \Phi^{-1}(1-\alpha) + nkp_0}$$

5) la puissance du test est

$$\pi = 1 - \beta = 1 - P[\text{Rejeter } H_1 | H_1 \text{ vraie}]$$

$$= P[\text{Rejeter } H_0 | H_1 \text{ vraie}]$$

$$= P[T_n > \gamma_\alpha | H_1 \text{ vraie}] = P\left[\frac{T_n - nkp_1}{\sqrt{nkp_1(1-p_1)}} > \frac{\gamma_\alpha - nkp_1}{\sqrt{nkp_1(1-p_1)}} \mid H_1 \right]$$

$$\text{d'où } \boxed{\pi = 1 - \Phi\left[\frac{\gamma_\alpha - nkp_1}{\sqrt{nkp_1(1-p_1)}} \right]}$$

Partie 3

7

1) La statistique du test de χ^2 est

$$\phi = \sum_i \frac{(k_i - np_i)^2}{np_i}$$

Ici on a naturellement 4 classes

$$C_0 (x_i=0) \quad k_0=5 \quad p_0 = P[X_i=0] = C_k^0 p_0^0 (1-p_0)^{k-0} = \left(1 - \frac{1}{2}\right)^3 = \boxed{\frac{1}{8}}$$

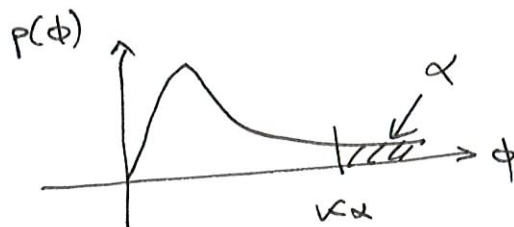
$$C_1 (x_i=1) \quad k_1=40 \quad p_1 = P[X_i=1] = C_k^1 p_0^1 (1-p_0)^{k-1} = 3 \left(\frac{1}{2}\right)^k = \boxed{\frac{3}{8}}$$

$$C_2 (x_i=2) \quad k_2=20 \quad p_2 = C_k^2 p_0^2 (1-p_0)^{k-2} = \boxed{\frac{3}{8}}$$

$$C_3 (x_i=3) \quad k_3=15 \quad p_3 = C_k^k p_0^k (1-p_0)^0 = \boxed{\frac{1}{8}}$$

$$\text{donc } \phi = \frac{(5-10)^2}{10} + \frac{(40-30)^2}{30} + \frac{(20-30)^2}{30} + \frac{(15-10)^2}{10} = \frac{25}{10} + \frac{100}{30} + \frac{100}{30} + \frac{25}{10} = \frac{150+200}{30} = \boxed{\frac{350}{30} \approx 11.67}$$

Sous H_0 , on sait que $\phi \sim \chi_3^2$



On rejette H_0 si $\phi > k_\alpha$ - Donc

$$\alpha = P[\phi > k_\alpha | H_0 \text{ vraie}] = \int_{k_\alpha}^{+\infty} f_{\chi_3^2}(u) du$$

$$\text{Pour } \alpha = 0,01, \text{ on a } P = 0,01 \Rightarrow \boxed{k_{0,01} = 11.345}$$

$$\text{Pour } \alpha = 0,05, \text{ on a } P = 0,05 \Rightarrow \boxed{k_{0,05} = 7.815}$$

$\phi > k_{0.01}$ ~~on~~ donc ^{on} rejette H_0 avec $\alpha = 0,01$

$\phi > k_{0.05}$ donc on rejette H_0 avec $\alpha = 0,05$

8

3) On a $k_{0,05} < k_{0,01}$ car α est la probabilité de rejeter H_0 sachant H_0 vraie - Quand $\alpha = 0,05 (> 0,01)$, on rejette H_0 plus souvent que pour $\alpha = 0,01$, donc le seuil associé à $\alpha = 0,05$ est inférieur à celui associé à $\alpha = 0,01$.

4) On ne peut calculer la puissance du test $\pi = P[\text{rejeter } H_0 \mid H_1 \text{ vraie}]$ car la loi de ϕ n'est pas connue sous H_1 .