CORRECTION TD5
STATISTIQUES INFÉRENTIELLES

Exercice 1

On peut modéliser l'expérience comme un 3-échantillon pour la loi géométrique de paramétrique p inconnu, $X \sim Geom(p)$. Ici on doit choisir entre $p_1 = 5/6$, qui correspond à D_1 et $p_2 = 1/6$, qui correspond à D_2. Posons $q_1 = 1 - p_1$ et $q_2 = 1 - p_2$.

Question 1 :

La probabilité sous l'hypothèse qu'on a tiré D_1 au tout début de l'expérience, d'avoir le 3-echantillon $(3, 5, 4)$ est :

$$P_{D1}(3,5,4) = P_X(3)P_X(5)P_X(4) = q_1^2 p_1 q_1^4 p_1 q_1^3 p_1$$
$$P_{D1}(3,5,4) = q_1^9 p_1^3$$

$$P_{D1}(3,5,4) \cong 5.742421.10^{-8}$$

Question 2 :

$$P_{D2}(3,5,4) = q_2^9 p_2^3$$
$$P_{D2}(3,5,4) \cong 8,972532.10^{-8}$$

Question 3 : A la suite de cette expérience on choisit le modèle qui est le plus vraisemblable. Ici $P_{D2}(3,5,4) > P_{D1}(3,5,4)$ donc on conclue que vraisemblablement le dé D_2 est à l'origine de cet échantillon.

Exercice 2

Question 1 :

Soit $x_1, ..., x_n$ une réalisation d'un échantillon distribué selon une loi normale de paramètre μ inconnu et $\sigma = 1$. On note $P_\mu(x_1...x_n)$ la probabilité de cette réalisation sous l'hypothèse que l'echantillon est i.i.d suivant $N(\mu, 1)$.

$$P_\mu(x_1...x_n) = \frac{1}{\sqrt{2\pi}^n} \prod_{i=1}^{n} e^{-\frac{1}{2}(x_i-\mu)^2}$$

$$P_\mu(x_1...x_n) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$\max\limits_{\mu\in\mathbb{R}} P_\mu(x_1...x_n)$ est atteint quand $\min\limits_{\mu\in\mathbb{R}} \sum_{i=1}^{n}(x_i - \mu)^2$ est atteint. Posons pour tout $\mu \in \mathbb{R}$, $f(\mu) = \sum_{i=1}^{n}(x_i - \mu)^2$; $f \in C^\infty(R)$ et convexe.

Pour tout $\mu \in \mathbb{R}$, $f'(\mu) = 2\left(\sum\limits_{i=1...n} x_i - n\mu\right)$ donc,

$$\forall \mu \in \mathbb{R}, \quad f'(\mu) = 0 \quad \Longleftrightarrow \quad \mu = \frac{\sum\limits_{i=1...n} x_i}{n}$$

On reconnaît que $\overline{x}_n := \frac{\sum\limits_{i=1...n} x_i}{n}$.

On a donc que f admet un minimum en $\mu = \overline{x}_n$ donc $P_\mu(x_1...x_n)$ admet un maximum en $\mu = \overline{x}_n$.

L'estimateur du maximum de vraisemblance de μ est $\hat{\mu} = \overline{x}_n$.

Question 2 :

On suppose que $\sigma > 0$.

$$P_{\sigma^2}(x_1...x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\prod_{i=1}^{n} e^{-\frac{1}{2}(x_i/\sigma)^2}$$

$$P_{\sigma^2}(x_1...x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i/\sigma)^2} > 0$$

Considérons la log vraisemblance,

$$g(\sigma^2) := lnP_{\sigma^2}(x_1...x_n) = -\frac{1}{2}\sum_{i=1}^{n}(x_i/\sigma)^2 - \frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi)$$

$g : \mathbb{R}_*^+ \to \mathbb{R}$ est $C^1(\mathbb{R}_*^+)$, pour tout $t \in \mathbb{R}_*^+$,

$$g^{'}(t) = \frac{1}{2}\sum_{i=1}^{n} x_i^2/t^2 - \frac{n}{2t}$$

$$\forall t > 0, \quad g^{'}(t) = 0 \quad \Longleftrightarrow \quad t = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

Posons $\hat{V} = \frac{\sum_{i=1}^{n} x_i^2}{n}$ et étudions les variations de g. Si $0 < t < \hat{V}$, alors g est strictement croissante ; si $t > \hat{V}$, g est strictement décroissante. Comme g est continue elle atteint son maximum en $\hat{V}$; il en est de même pour la vraisemblance, donc l'estimateur du maximum de vraisemblance est $\hat{V}$.

Par ailleurs,

$$E_{\sigma^2}[\hat{V}] = \frac{1}{n}\sum_{i=1}^{n} E_{\sigma^2}[X_i^2] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2$$

$$E_{\sigma^2}[\hat{V}] = \sigma^2$$

Donc,

$$B(\hat{V}) = E_{\sigma^2}[\hat{V}] - \sigma^2 = 0$$

Exercice 3

La vraisemblance de la réalisation est,

$$P_\delta(x_1...x_n) = e^{-\sum\limits_{i=1...n}(x_i-\delta)} 1[\min_{i=1...n} x_i \geq \delta]$$

Posons, $f(\delta) = e^{-\sum\limits_{i=1...n}(x_i-\delta)} = e^{n\delta}e^{-\sum\limits_{i=1...n} x_i}$ pour $\delta \leq \min\limits_{i=1...n} x_i$; elle est croissante donc son maximum est atteint en $\min\limits_{i=1...n} x_i$. Par ailleurs, $e^{n\min\limits_{i=1...n} x_i}e^{-\sum\limits_{i=1...n} x_i} > 0$ donc le maximum de vraisemblance est atteint en $\hat{\delta} = \min\limits_{i=1...n} x_i$.

Exercice 4

On rappelle que la fontion de répartition de la loi de Pareto de paramètre $\alpha > 0$ est donnée par, pour $x < 1$,

$$F_\alpha(t) = 0$$

pour $x \geq 1$,

$$F_\alpha(t) = 1 - t^{-\alpha}$$

F_α est C^1 par morceaux, donc elle admet une densité de probabilité (par exemple) donnée par, si $x < 1$,

$$f_\alpha(x) = 0$$

et si $x \geq 1$,

$$f_\alpha(x) = \alpha t^{-(\alpha+1)}$$

Reparamétrons cette densité en $\beta = 1/\alpha$, pour $x < 1$,

$$f_\beta(x) = 0$$

pour $x \geq 1$,

$$f_\beta(x) = \frac{1}{\beta} t^{-(1+\frac{1}{\beta})} = \frac{1}{\beta} e^{-(1+\frac{1}{\beta})\ln(t)}$$

La vraisemblance est donnée par,

$$P_\beta(x_1 ... x_n) = \frac{1}{\beta^n} e^{-(1+\frac{1}{\beta}) \sum\limits_{i=1...n} \ln x_i} 1[\min_{i=1...n} x_i \geq 1]$$

Posons $f : \mathbb{R}_*^+ \to R$, pour tout $\beta \in \mathbb{R}_*^+$, $f(\beta) = P_\beta(x_1 ... x_n)$; par ailleurs $f \in C^\infty(\mathbb{R}_*^+)$. Si $\min\limits_{i=1...n} x_i < 1$ tout nombre dans $\mathbb{R}_*^+$ maximise la vraisemblance, si $\min\limits_{i=1...n} x_i \geq 1$,

$$\forall t \in \mathbb{R}_*^+, f^{'}(t) \geq 0 \quad \Longleftrightarrow \quad -\frac{n}{\beta^{n+1}} + \frac{\sum\limits_{i=1...n} \ln x_i}{\beta^{n+2}} = 0$$

$$\forall t \in \mathbb{R}_*^+, f^{'}(t) \geq 0 \quad \Longleftrightarrow \quad \beta \leq \frac{\sum\limits_{i=1...n} \ln x_i}{n}$$

Comme f est C^∞ elle est en particulier continue, donc elle atteint son maximum en $\frac{\sum_{i=1...n} \ln x_i}{n}$.

Exercice 5

Question 1

Pour tout $0 < \theta < 1$,

$$L(\theta) = p(1)^1 997 p(2)^3 2 p(3)^9 06 p(4)^9 04$$

$$l(\theta) = 1997 \ln p(1) + 32 \ln p(2) + 906 \ln p(3) + 904 \ln p(4)$$

$$l(\theta) = 1997 \ln(2+\theta) + 32 \ln(\theta) + 1810 \ln(1-\theta) - 7646 \ln 2$$

Question 3

Supposons que n_1, n_2, n_3, n_4 sont strictement positifs.

Il existe une constante C telle que pour tout $\theta \in]0,1[$,

$$l(\theta) = n_1 \ln(2+\theta) + n_2 \ln(\theta) + (n_3 + n_4) \ln(1-\theta) + C$$

Posons $n_5 = n_3 + n_4$.

$l \in C^1(]0,1[)$, pour tout $\theta \in]0,1[$,

$$l^{'}(\theta) = \frac{n_1}{2+\theta} + \frac{n_2}{\theta} - \frac{n_5}{1-\theta}$$

$$\forall \theta \in]0,1[\quad l^{'}(\theta) \geq 0 \qquad \Longleftrightarrow \qquad n_1\theta(1-\theta) + n_2(2+\theta)(1-\theta) - n_5(2+\theta)\theta \geq 0$$

Posons, pour $\theta \in \mathbb{R}$, $P(\theta) = n_1\theta(1-\theta) + n_2(2+\theta)(1-\theta) - n_5(2+\theta)\theta$, c'est un polynome du second degré.

Pour $\theta \in \mathbb{R}$,

$$P(\theta) = n_1(\theta - \theta^2) + n_2(2 - 2\theta + \theta - \theta^2) - n_5(2\theta + \theta^2) = -n\theta^2 + (n_1 - n_2 - 2n_5)\theta + 2n_2$$

$\Delta = (n_1 - n_2 - 2n_5)^2 + 8nn_2 \geq 0$. Donc une racine de P est $\theta_1 = \frac{-(n_1-n_2-2n_5)-\sqrt{\Delta}}{2n}$ et l'autre est $\theta_2 = \frac{-(n_1-n_2-2n_5)+\sqrt{\Delta}}{2n}$; Par ailleurs, $\theta_2 \geq \theta_1$.

Pour $\theta \in]0,1[$ le signe de $l^{'}$ est le signe de P ; or P est positif en ses racines et négatif en dehors.

On remarque que $\lim\limits_{\theta \to 0} l^{'}(\theta) = +\infty$ et que $\lim\limits_{\theta \to 1} l^{'}(\theta) = -\infty$ donc le signe de P sur $]0,1[$ doit etre positif puis negatif et ceci n'est possible que si $\theta_2 \in]0,1[$ et $\theta_1 < 0$. Ansi on trouve que pour tout $\theta \in]0, \theta_2[$ $l'(\theta) > 0$ et pour $\theta \in]\theta_2, 1[$ $l'(\theta) < 0$ donc

$$\hat{\theta} = \theta_2 = \frac{-(n_1 - n_2 - 2n_5) + \sqrt{\Delta}}{2n}$$