Statistiques fondamentales **Université de Paris – M1**

Partiel 2022-2023 – Corrigé

Exercice 1 - Questions

1. On peut définir une loi $\chi^2(n)$ par sa densité
$$t^{\frac{n}{2}-1}e^{-\frac{t}{2}}\frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma(n/2)}\mathbb{1}_{t>0},$$
et la loi de Student $\mathcal{T}(n)$ via $\mathcal{T}(n)\sim\frac{N}{\sqrt{S}}$, où $N\sim\mathcal{N}(0,1)$, $S\sim\chi^2(n)/n$, et N, S sont indépendantes.
2. Erreur de première espèce : $\sup_{\theta\in\Theta_0}P_\theta(T=1)$. Puissance (uniforme) : $\inf_{\theta\in\Theta_1}P_\theta(T=1)$. Théorème de Neyman Pearson : si $\Theta_j=\{\theta_j\}$ et si le test du rapport de vraisemblance est de niveau exact (α), alors ce test est le plus puissant possible (parmi les tests de niveau α).
3. Comme $\hat{\sigma}_n^2/\sigma^2\sim\chi^2(n-1)/(n-1)$ est indépendante de $\bar{x}_n$ (application directe de Cochran), et que $\sqrt{n}(\bar{x}_n-\mu)/\sigma\sim\mathcal{N}(0,1)$, on a que $\sqrt{n}(\bar{x}_n-\mu)/\hat{\sigma}_n\sim\mathcal{T}(n-1)$. En notant q le $1-(\alpha/2)$- quantile d'une telle loi, $[\bar{x}_n\pm\hat{\sigma_n}q/\sqrt{n}]$ est un IC de niveau de confiance $1-\alpha$.

Exercice 2 - Un estimateur de covariance

1. Modèle : $((\mathbb{R}^2)^n,\mathcal{B}((\mathbb{R}^2)^n),(\mathcal{N}(0,\Sigma_\theta)^{\otimes n})_{\theta\in]0_2,\mathrm{arcsh}(1)[})$.
2. Si $(U,V)^T\sim\mathcal{N}(0,I_2)$, alors $(U,\mathrm{sh}(\theta)U+\sqrt{1-\mathrm{sh}^2(\theta)}V)^T=A(U,V)^T$, avec
$$A=\begin{pmatrix}1 & 0\\ \mathrm{sh}(\theta) & \sqrt{1-\mathrm{sh}^2(\theta)}\end{pmatrix},$$
est un vecteur Gaussien, de moyenne $A(0,0)^T=(0,0)^T$, et de covariance
$$AA^T=\begin{pmatrix}1 & \mathrm{sh}(\theta)\\ sh(\theta) & 1\end{pmatrix}=\Sigma_\theta.$$
3. On a $\mathbb{E}(\chi^2(1))=\mathrm{Var}(\mathcal{N}(0,1))=1$. Par ailleurs
$$\begin{aligned}\mathbb{E}((\chi^2(1))^2)&=\int_0^{+\infty}t^{\frac{3}{2}}e^{-\frac{t}{2}}\frac{1}{\sqrt{2}\Gamma(1/2)}dt\\&=\frac{\Gamma(5/2)}{(1/2)^{(5/2)}}\frac{1}{\sqrt{2}\Gamma(1/2)}\\&=\frac{3\times 2^{(5/2)}}{4\sqrt{2}}=3.\end{aligned}$$
On en déduit $\mathrm{Var}(\chi^2(1))=2$.

4. En utilisant la question 2, on peut écrire

$$\begin{aligned}\mathrm{Var}\left(\mathrm{sh}(\theta)U^2+\sqrt{1-\mathrm{sh}^2(\theta)}UV\right) &= \mathrm{sh}^2(\theta)\mathrm{Var}(U^2)+(1-\mathrm{sh}^2(\theta))\mathrm{Var}(UV)\\ &\quad +2\mathrm{sh}(\theta)\sqrt{1-\mathrm{sh}^2(\theta)}\mathrm{Cov}(U^2,UV),\end{aligned}$$

avec

$$\begin{aligned}\mathrm{Var}(U^2) &= 2\\ \mathrm{Var}(UV) &= \mathbb{E}((UV)^2)-(\mathbb{E}(UV))^2=\mathbb{E}(U^2)\mathbb{E}(V^2)=1\\ \mathrm{Cov}(U^2,UV) &= \mathbb{E}(U^3V)-\mathbb{E}(U^2)\mathbb{E}(UV)=0.\end{aligned}$$

On en déduit

$$\mathrm{Var}(XY)=2\mathrm{sh}^2(\theta)+(1-\mathrm{sh}^2(\theta))=1+\mathrm{sh}^2(\theta).$$

5. Comme $E_\theta(XY)=\mathrm{sh}(\theta)$, un estimateur par la méthode des moments est

$$\hat\theta=\mathrm{arcsh}\left(\overline{XY}\right)=\mathrm{arcsh}\left(\frac{1}{n}\sum_{i=1}^n X_iY_i\right).$$

Comme $E_\theta(|XY|)\leq\sqrt{E_\theta(XY)^2}<+\infty$, la loi des grands nombres donne

$$\overline{XY}\underset{n\to+\infty}{\longrightarrow}\mathrm{sh}(\theta),$$

en probabilité. arcsh étant continue, $\hat\theta$ est consistant.

6. Comme $\mathrm{Var}_\theta(XY)=1+\mathrm{sh}^2(\theta)<+\infty$, le théorème central limite donne

$$\sqrt{n}\left(\overline{XY}-\mathrm{sh}(\theta)\right)\rightsquigarrow\mathcal{N}(0,1+\mathrm{sh}^2(\theta)).$$

Comme arcsh est différentiable sur $\mathbb{R}$, la méthode Δ donne

$$\sqrt{n}\left(\mathrm{arcsh}(\overline{XY})-\theta\right)\rightsquigarrow\mathrm{arcsh}'(\mathrm{sh}(\theta))\mathcal{N}(0,1+\mathrm{sh}^2(\theta)).$$

Or, $\mathrm{arcsh}'(\mathrm{sh}(\theta))=\frac{1}{\mathrm{sh}'(\theta)}=\frac{1}{\sqrt{1+\mathrm{sh}^2(\theta)}}$. On en déduit alors

$$\sqrt{n}\left(\hat\theta-\theta\right)\rightsquigarrow\mathcal{N}(0,1).$$

7. Si q est le quantile d'ordre $1-(\alpha/2)$ d'une loi Normale standard, $[\hat\theta\pm q/\sqrt{n}]$ est un intervalle de niveau de confiance $1-\alpha$.

Exercice 3 - Pêche dans l'étang d'Orsay

1. On tire au hasard n éléments parmi N, sans remise et sans ordre. Le nombre de tels tirages possibles est $\binom{N}{n}$, et on suppose qu'il y a équiprobablité sur les issues. En notant A_k l'évènement "Pascal a pêché k carpes", on a
$$|A_k| = \binom{Np}{k}\binom{N(1-p)}{n-k},$$
correspondant aux choix de k éléments parmi les Np carpes et $n-k$ parmi les $N(1-p)$ autres poissons. On en déduit
$$\mathbb{P}_p\left(S=k\right) = \frac{|A_k|}{\binom{N}{n}} = \frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}.$$

2. Si $(X_1,\ldots,X_n)$ est échangeable, $(Y_1,\ldots,Y_n)$ l'est aussi. Alors, pour tout $j \in [\![1,n]\!]$, $E_p(Y_j) = E_p(Y_1)$ (choisir la permutation qui échange 1 et j). Comme $Y_1 \sim \mathcal{B}(p)$, $E_p(Y_1) = p$, et $E_p(S) = np$. On en déduit que $\hat{p} = S/n$ est un estimateur sans biais de p.

3. Y_1Y_2 vaut 1 si les deux premiers poissons pêchés sont des carpes, 0 sinon. On en déduit que
$$Y_1Y_2 \sim \mathcal{B}\left(\frac{\binom{Np}{2}}{\binom{N}{2}}\right) \sim \mathcal{B}\left(\frac{Np(Np-1)}{N(N-1)}\right).$$
Comme $(Y_1,\ldots,Y_n)$ est échangeable, on a, pour tous $1 \leq i \neq j \leq n$,
$$\begin{aligned}\mathrm{Cov}_p(Y_i,Y_j) = \mathrm{Cov}_p(Y_1,Y_2) &= \frac{p(Np-1)}{N-1} - p^2\\ &= \frac{Np^2 - p - p^2N + p^2}{N-1} = \frac{p(p-1)}{N-1}.\end{aligned}$$
On en déduit
$$\begin{aligned}\mathrm{Var}_p(S) &= \sum_{i=1}^{n}\mathrm{Var}_p(Y_i) + 2\sum_{1\leq i<j\leq n}\mathrm{Cov}_p(Y_i,Y_j)\\ &= n\mathrm{Var}_p(Y_1) + \frac{np(n-1)(p-1)}{N-1}\\ &= np(1-p)\left(1 - \frac{n-1}{N-1}\right).\end{aligned}$$

4. L'inégalité de Bienaymé-Cebicev (ainsi que $p(1-p) \leq 1/4$) donne, pour $t > 0$,
$$\mathbb{P}_p\left(\hat{p} > p+t\right) \leq \frac{\left(1 - \frac{n-1}{N-1}\right)}{4nt^2}.$$

En posant

$$t_\alpha = \sqrt{\frac{\left(1 - \frac{n-1}{N-1}\right)}{4n\alpha}},$$

un intervalle de niveau de confiance $1-\alpha$ est donné par $[\hat{p} - t_\alpha, +\infty[$.

5. Soit $k \in [\![0, n]\!]$. On a

$$\begin{aligned}
&\mathbb{P}_p(S = k) \\
&= \binom{n}{k} \frac{Np(Np-1)\dots(Np-k+1) \times N(1-p)(N(1-p)-1)\dots(N(1-p)-(n-k)+1)}{N(N-1)\dots(N-n+1)} \\
&= \binom{n}{k} \frac{(Np)^k (N(1-p))^{n-k}}{N^n} \times \\
&\quad \frac{\left(1-\frac{1}{Np}\right)\left(1-\frac{2}{Np}\right)\dots\left(1-\frac{k-1}{Np}\right) \times \left(1-\frac{1}{N(1-p)}\right)\left(1-\frac{2}{N(1-p)}\right)\dots\left(1-\frac{n-k-1}{N(1-p)}\right)}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{n-1}{N}\right)} \\
&\underset{N\to+\infty}{\longrightarrow} \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

6. Si $Z \sim \mathcal{B}(n,p) \sim \sum_{i=1}^n Z_i$, où les Z_i sont i.i.d. $\sim \mathcal{B}(p)$, l'inégalité de Hoeffding donne

$$\mathbb{P}_p\left(\frac{Z}{n} \geq p + t\right) \leq e^{-2nt^2}.$$

En posant $t_\alpha = \sqrt{\frac{\log(1/\alpha)}{2n}}$, comme $S \rightsquigarrow Z$ lorsque $N \to +\infty$, on en déduit que $[(S/n) - t_\alpha, +\infty[$ est un intervalle de niveau de confiance asymptotique $1-\alpha$ (meilleur que le précédent lorsque $\alpha \to 0$).

7. On souhaite bâtir un test de niveau α pour les hypothèses

$$\begin{cases} H_0 & : \quad p \leq 0.1 \\ H_1 & : \quad p > 0.1. \end{cases}$$

Si on prend $T = \mathbb{1}_{(S/n) \geq 0.1 + t_\alpha}$, avec t_α comme en $4.$ ou $6.$, on a, pour tout $p \leq 0.1$,

$$\mathbb{P}_p((S/n) \geq 0.1 + t_\alpha) \leq \mathbb{P}_p((S/n) \geq p + t_\alpha) \leq \alpha.$$

T est donc un test de niveau α.

8. En prenant t_α comme en $4.$, pour $\alpha = 0.1$ et $n = 9$, on a

$$t_\alpha = \sqrt{\frac{1 - \frac{n-1}{N-1}}{4n\alpha}} \geq \sqrt{\frac{9/10}{36/10}} = \frac{1}{2}.$$

On doit alors avoir $S/9 > 0.1 + (1/2)$, ce qui équivaut à $S > 5.4$. Pascal doit donc pêcher plus de 6 carpes pour être sûr à 90% de renouveler son permis de pêche à bon escient.