Final SY02 Printemps 2020

— Répondre de manière manuscrite sur des feuilles au format A4
— Veiller à ajouter les nom, prénom et signature en haut de la première page de la copie.
— La qualité de la présentation sera prise en compte dans la notation.

Exercice 1. (2 points)

On considère un échantillon iid $X_1, \dots, X_n$ de loi parente $\mathcal{N}(\mu, 1)$.

1. $\boxed{\text{0.5 pts}}$ On considère l'estimateur $\widehat{\mu}_1 = 1$. Quel est le bias et la variance de cet estimateur ?
2. $\boxed{\text{0.5 pts}}$ On suppose maintenant que $\widehat{\mu}_2 = X_1$. Quel est le biais et la variance de cet estimateur ?
3. $\boxed{\text{0.5 pts}}$ On suppose enfin que $\widehat{\mu}_3 = \overline{X}$. Quel est le biais et la variance de cet estimateur ?
4. $\boxed{\text{0.5 pts}}$ Dire en une phrase lequels de ces estimateurs est le meilleur et pourquoi.

Exercice 2. (5 points)

On considère l'échantillon suivant

$$-3.97, \quad 2.22, \quad 4.35, \quad 4.76.$$

1. $\boxed{\text{5 pts}}$ Tester au niveau $\alpha^* = 0.05$ si cet échantillon suit une loi normale d'espérance 2 et d'écart-type 3. On calculera la borne de la région critique ainsi que le degré de signification.

Exercice 3. (6 points)

On modélise le résultat d'un sondage par le tirage d'un échantillon de longueur 1 d'une loi binomiale $\mathcal{B}(n, p)$. On souhaite réaliser le test T_1 suivant sur le paramètre p de la loi binomiale :

$$\begin{cases} H_0 : & p = p_0 = 0.49, \\ H_1 : & p = p_1 = 0.51. \end{cases}$$

On sait que la région critique du test optimal s'écrit,

$$W = \{x > c\},$$

avec c une constante à déterminer.

1. $\boxed{\text{1.5 pts}}$ En utilisant l'approximation normale (sans la correction de continuité), trouver une expression du seuil c en fonction de n, α^* et p_0.
2. $\boxed{\text{1 pt}}$ Quel est le seuil c et le résultat du test T_1 au niveau $\alpha^* = 0.05$ si on suppose que le sondage est réalisé sur 100 personnes et qu'on observe $x = 55$.

On souhaite trouver n tel que le risque de seconde espèce vaut lui aussi α^*.

3. $\boxed{\text{1 pt}}$ Trouver une seconde expression de c en fonction cette fois de n, $\beta = \alpha^*$ et p_1.
4. $\boxed{\text{1 pt}}$ Déduire des deux questions précédentes une égalité portant sur n, α^*, p_0 et p_1.
5. $\boxed{\text{1.5 pts}}$ En déduire l'expression de n en fonction de α^*, p_0 et p_1 et appliquer numériquement.

Exercice 4. (13 points)

Pour faire une tarte aux cerises, on pioche des fruits au hasard dans un panier et on les ouvre. Si le fruit est véreux (colonisé par *drosophila suzukii*), il est jugé impropre à la consommation et laissé de côté ; dans le cas contraire, on l'utilise. Une fois le nombre désiré r de fruits non colonisés atteint, on compte également le nombre k de fruits laissés de côté, afin d'estimer la probabilité p qu'un fruit ne soit pas gâté.

On admettra que la distribution du nombre de fruits K gâtés suit une loi binomiale négative : sa fonction de probabilité est définie, pour tout $k = 0, 1, 2, \ldots$, par

$$\Pr(K = k; r, p) = C_{k+r-1}^{k} p^r (1-p)^k.$$

On supposera que $p \in]0; 1[$. Le paramètre r est fixé par la recette et donc connu.

On admettra que la variable aléatoire K admet pour espérance et variance

$$\mathbb{E}[K] = \frac{r(1-p)}{p}, \quad \mathrm{Var}(K) = \frac{r(1-p)}{p^2}.$$

1. $\boxed{\text{1 pt}}$ Calculer un estimateur $\widehat{p}_m$ de p par la méthode des moments. Est-il convergent ? Justifier.
2. $\boxed{\text{1.5 pts}}$ L'estimateur $\widehat{p}_m$ est-il sans biais ? Justifier.
3. $\boxed{\text{1 pt}}$ Calculer la fonction de vraisemblance du paramètre p étant donné un couple de nombres (r, k) de fruits non gâtés et gâtés (on n'a fait qu'une seule tarte), puis la log-vraisemblance.
4. $\boxed{\text{1.5 pts}}$ Calculer l'estimateur du maximum de vraisemblance $\widehat{p}$ du paramètre p. On justifiera toutes les étapes du raisonnement.
5. $\boxed{\text{1 pt}}$ Comparer cet estimateur à celui obtenu par la méthode des moments, en justifiant.
6. $\boxed{\text{2 pts}}$ Calculer l'information de Fisher apportée par l'échantillon relativement au paramètre p. En déduire la loi asymptotique d'une fonction de p que l'on précisera.
7. $\boxed{\text{2 pts}}$ En déduire une fonction pivotale approchée, puis un intervalle de confiance bilatéral approché de niveau $1 - \alpha$, pour le paramètre p. On fera le moins d'approximations possible.
8. $\boxed{\text{1.5 pts}}$ Calculer la réalisation de cet intervalle de confiance, avec les données suivantes, pour $\alpha = 0.05$: $r = 117, k = 872$.
9. $\boxed{\text{1.5 pts}}$ À partir de cerises cueillies sur un autre arbre, on a compté les nombres suivants de cerises utilisées r' et laissées de côté car colonisées k' : $r' = 87, k' = 150$. Proposer une stratégie approchée pour déterminer si la proportion de cerises non colonisées par *drosophila suzukii* est significativement différente d'un arbre à l'autre. Qu'en est-il avec les données de l'exercice ?

Exercice 5. (4 points)

Soit X_1, X_2 un échantillon de longueur 2 de loi parente la loi continue uniforme sur l'intervalle $[\theta, \theta + 1]$. On désire effectuer le test suivant :

$$\begin{cases} H_0 : & \theta = 0, \\ H_1 : & \theta > 0. \end{cases}$$

Soit W_1 une région critique définie par

$$W_1 = \big\{(x_1, x_2) \in \mathbb{R}^2,\ x_1 > 0.9\big\}.$$

1. $\boxed{\text{1 pt}}$ Quelle est le risque de première espèce associé à la région critique W_1 ?

On introduit une seconde région critique définie comme suit

$$W_2 = \big\{(x_1, x_2) \in \mathbb{R}^2,\ x_1 + x_2 = c\big\}.$$

Il n'est pas difficile de montrer que la variable aléatoire $X_1 + X_2$ suit une loi dite triangulaire définie par la densité f_θ suivante :

$$f_\theta(x) = \begin{cases} 4(x - \theta) & \text{si } x \in \big[\theta, \theta + \frac{1}{2}\big] \\ 4(\theta + 1 - x) & \text{si } x \in \big[\theta + \frac{1}{2}, \theta + 1\big] \\ 0 & \text{sinon.} \end{cases}$$

2. $\boxed{\text{1 pt}}$ Tracer la densité en faisant figurer θ et $\theta + 1$ en abscisse et la valeur du maximum en ordonnée.
3. $\boxed{\text{2 pts}}$ Déterminer la constante c de telle manière que le risque de première espèce de W_2 soit égal à 0.08 (une démonstration géométrique sera acceptée).