Corrigé Rattrapage Stats fév 22

Exo 1 :

Justifier si les propositions suivantes sont vraies ou fausses :

1. Dans le cas le plus général, l'IDC à 95% pour une proportion est plus large qu'à 99%.

   Faux, IDC $= f_n \pm t\sqrt{\frac{f_n(1-f_n)}{n}}$. Pour un IDC à 95% la valeur de t donnée en général par la table de la loi normale est $t = 1.96$ alors que pour 99% elle est de $t = 2.58$ L'IDC sera plus large à 99%.

2. Lorsqu'on applique un test d'indépendance du khi-deux, le risque de première espèce est de décider que les variables sont liées alors qu'elles sont indépendantes.

   Vrai, dans ce genre de test, (H_0) correspond à l'indépendance.
   $\alpha =$ Risque de 1ère espèce= Proba. de valider (H_1) à tort.

3. On rejette l'hypothèse (H_1) lorsque la p-valeur est plus petite que le risque de première espèce α.

   Faux, la p-valeur peut être vue comme "la proba. que (H_0) soit vraie étnt donné les valeurs de mon échantillon". Si cette valeur est trop petite (inférieure à α), on rejette (H_0).

Exo 2 :

Soit $\theta \in]0,1[$ un paramètre inconnu et X une variable aléatoire dont la densité est donnée par :

$$f_X(x) = \begin{cases} (1-\theta) + \dfrac{\theta}{2\sqrt{x}} & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

On dispose de $(X_1, \dots, X_n)$ un n-échantillon de X. On note $\overline{X}$ la moyenne empirique de X.

1. Montrer que $E(X) = \frac{1}{2} - \frac{\theta}{6}$.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^1 \left((1-\theta)x + \frac{\theta x}{2\sqrt{x}}\right)dx$$

$$= \int_0^1 \left((1-\theta)x + \frac{\theta\sqrt{x}}{2}\right)dx = \left[(1-\theta)\frac{x^2}{2} + \frac{\theta}{3}x\sqrt{x}\right]_0^1$$

$$= (1-\theta)\frac{1}{2} + \frac{\theta}{3} = \frac{1}{2} - \frac{\theta}{6}.$$

On admettra que $Var(X) = \frac{1}{12} + \frac{\theta}{30} - \frac{\theta^2}{36}$.

2. On considère $T_1 = \overline{X}$ comme un premier estimateur de θ.
   Calculer son biais et son risque quadratique.

$$E(T_1) = E(X) = \frac{1}{2} - \frac{\theta}{6} \quad \Longrightarrow b(T_1) = E(T_1) - \theta = \frac{1}{2} - \frac{7\theta}{6}$$

$$R_\theta(T_1) = Var(T_1) + b(T_1)^2 = Var(\overline{X}) + \left(\frac{1}{2} - \frac{7\theta}{6}\right)^2$$

$$R_\theta(T_1) = \frac{Var(X)}{n} + \left(\frac{1}{2} - \frac{7\theta}{6}\right)^2$$

$$R_\theta(T_1) = \frac{\frac{1}{12} + \frac{\theta}{30} - \frac{\theta^2}{36}}{n} + \left(\frac{1}{2} - \frac{7\theta}{6}\right)^2$$

3. T_1 est-il un estimateur convergent ?

$$\lim_{n \to +\infty} R_\theta(T_1) = \left(\frac{1}{2} - \frac{7\theta}{6}\right)^2 \neq 0 \quad \text{sauf dans le cas } \theta = \frac{3}{7}.$$

T_1 n'est pas convergent sauf dans le cas $\theta = \frac{3}{7}$

4. Déterminer les réels a et b pour que l'estimateur $T_2 = a + bT_1 = a + b\overline{X}$ soit un estimateur sans bias de θ.
   Pour que T_2 soit sans biais, il faut avoir $E(T_2) = \theta$, ce qui équivaut à :

$$\theta = E(a + b\overline{X}) = a + bE(X) = a + b\left(\frac{1}{2} - \frac{\theta}{6}\right) = a + \frac{b}{2} - b\frac{\theta}{6}$$

$$\Longleftrightarrow \begin{cases} a + \frac{b}{2} & = 0 \\ -\frac{b}{6} & = 1 \end{cases} \Longleftrightarrow \begin{cases} a & = 3 \\ b & = -6 \end{cases}$$

5. Calculer le risque quadratique de T_2 et conclure quant à sa convergence.

$$R_\theta(T_2) = Var(T_2) + 0 = Var(3 - 6\overline{X})$$

$$R_\theta(T_2) = Var(-6\overline{X}) = 36Var(\overline{X}) = 36\frac{Var(X)}{n}$$

$$R_\theta(T_2) = 36\frac{\frac{1}{12} + \frac{\theta}{30} - \frac{\theta^2}{36}}{n} = \frac{3}{n} + \frac{6\theta}{5n} - \frac{\theta^2}{n}$$

$$\lim_{n \to +\infty} R_\theta(T_2) = 0 \quad \Longrightarrow \text{Estimateur convergent.}$$

Exo 3 :

Voulant évaluer rapidement les résultats obtenus par ses 200 étudiants lors d'un partiel, un professeur décide de corriger quelques copies tirées au hasard.
Il admet par ailleurs que les notes de ses élèves suivent une loi normale $\mathcal{N}\left(\mu, \sigma^2\right)$ de variance $\sigma^2 = 4$.
Le professeur corrige un échantillon de $n = 7$ copies et trouve une moyenne de $\overline{x} = 11$.

1. Donner un intervalle de confiance à 95% de la moyenne des 200 copies ?

   $$\overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 11 \pm 1.96\frac{2}{\sqrt{11}} = [9.81 \ \ ; \ \ 12.18]$$

2. Combien de copies le professeur doit-il corriger s'il veut situer la moyenne générale de ses
   élèves dans un intervalle de confiance d'amplitude 2, avec le même niveau de confiance de 95% ?

   Il faudrait avoir : $1.96\frac{\sigma}{\sqrt{n}} = 1 \quad \Longleftrightarrow \quad \sqrt{n} = 1.96\sigma \quad \Longleftrightarrow \quad n = 15.4$

   Il faut corriger 16 copies.

3. En trouvant une moyenne égale à $\overline{x} = 11$, combien de copies le professeur devrait-il corriger pour pouvoir dire, avec un risque de 1%, que la moyenne de tous les élèves est supérieure à 10 ?

   Même calcul que la question précédente mais avec 99%, il faut prendre $t = 2.58$ et non pas 1.96.

   On trouve donc : $\quad n = (2.58 \times 2)^2 = 26.6.$

   Il faut corriger 27 copies.

Exo 4 :

Une machine fabrique des pièces dont la longueur suit une loi normale $\mathcal{N}\left(\mu, \sigma^2\right)$,où μ est inconnu et $\sigma = 2$.
Lorsque la machine fonctionne correctement, elle produit des pièces de 100 cm de long.
Lorsqu'elle est déréglée, elle peut aussi bien produire des pièces trop longues que trop courtes.
On désire construire un test pour savoir si elle est déréglée, et s'il faut donc faire intervenir une société spécialisée pour la réparer.

1. Enoncer les 2 hypothèses et expliciter les risques de 1ère et de 2ème espèce.

   $$\begin{cases} (H_0) & \mu = \mu_0 = 100 \\ (H_1) & \mu = \mu_1 \neq 100 \end{cases}$$

α = Proba. de décider de réparer la machine alors qu'elle fonctionne bien.
β = Proba. de ne pas réparer la machine alors qu'elle est déréglée.

2. Quelle sera la variable de décision et quelle sera la loi utilisée ?

   Var de décision : $\overline{X}$ et loi normale.

3. Faire une représentation graphique montrant la région critique et le risque de 1ère espèce.

4. Avec un premier échantillon de taille $n = 10$, on obtient une moyenne $\overline{x} = 99$ cm. Doit on rejeter H_0 ?

   Avec un risque de $\alpha = 5\%$, les seuils sont calculés de la manière suivante :

   $$C_1 = \mu_0 - 1.96\frac{\sigma}{\sqrt{n}} = 100 - 1.96\frac{2}{\sqrt{10}} = 100 - 1.23 = 98.77$$

   $$C_2 = \mu_0 + 1.96\frac{\sigma}{\sqrt{n}} = 100 + 1.96\frac{2}{\sqrt{10}} = 100 + 1.23 = 101.23$$

   $\overline{x} = 99 \in [C_1, C_2] \implies$ On ne doit pas rejeter (H_0), on le valide.

5. Avec un deuxième échantillon de taille $n = 50$, on obtient une moyenne $\overline{x} = 99$ cm. Doit on rejeter H_0 ?

   Avec un risque de $\alpha = 5\%$, les seuils sont calculés de la manière suivante :

   $$C_1 = \mu_0 - 1.96\frac{\sigma}{\sqrt{n}} = 100 - 1.96\frac{2}{\sqrt{50}} = 100 - 0.55 = 99.45$$

   $$C_2 = \mu_0 + 1.96\frac{\sigma}{\sqrt{n}} = 100 + 1.96\frac{2}{\sqrt{50}} = 100 + 0.55 = 100.55$$

   $\overline{x} = 99 \notin [C_1, C_2] \implies$ On doit rejeter (H_0), on valide (H_1).

Exo 5 :

Le tableau ci-dessous contient les nombres $N_{i,obs}$ d'apparition des entiers x_i (de 0 à 9) dans les 10000 premières décimales du nombre π.

x_i	0	1	2	3	4	5	6	7	8	9
$N_{i,obs}$	968	1025	1021	974	1014	1045	1021	970	948	1014

Nous souhaitons tester l'hypothèse d'une répartition uniforme de ces entiers.

1. Quel est le nom du test à effectuer, et quelles sont les hypothèses ?
   Test du khi 2 d'adéquation à une loi discrète.
   $\begin{cases} (H_0) & \text{L'échantillon suit la loi} \\ (H_1) & \text{L'échantillon ne suit pas la loi} \end{cases}$

2. Compléter le tableau suivant permettant de répondre à la question précédente.

x_i	0	1	2	3	4	5	6	7	8	9
Effectif observé $N_{i,obs}$	968	1025	1021	974	1014	1045	1021	970	948	1014
Effectif théorique $N_{i,theo}$	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
Distance	1.024	0.625	0.441	0.676	0.196	2.025	0.441	0.9	2.704	0.196

3. Donner la formule permettant de calculer la distance totale ainsi que la loi suivie sous (H_0).

$$d_i = \frac{(N_{i,obs} - N_{i,theo})^2}{N_{i,theo}}$$

4. Donner la conclusion de ce test avec un risque de $\alpha = 5\%$.

   $D = \sum d_i = 9.228$
   Sous (H_0), D suit la loi du khi-deux à $10 - 1 = 9$ d.d.l.
   Le seuil donné par la table à 5% est de $C = 16.919$.
   $D < C \quad \Longrightarrow$ On ne peut pas rejeter (H_0).
   La répartition des chiffres des 10000 décimales de π est bien uniforme