STT1000 - Exercices # 5

Automne 2021

Exercice 1 – ***(Loi exponentielle)***
On considère n observations $\{y_1, \cdots, y_n\}$ i.i.d. suivant une loi exponentielle $\mathcal{E}(\lambda)$, de moyenne $1/\lambda$. On pose $\theta = 1/\lambda$. Soient

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ et } \widetilde{\theta} = \frac{1}{n+1}\sum_{i=1}^{n} y_i$$

Comparer l'erreur quadratique moyenne de $\widehat{\theta}$ et de $\widetilde{\theta}$

Exercice 2 – ***(Loi normale)***
On considère n observations $\{y_1, \cdots, y_n\}$ i.i.d. suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$, avec n relativement grand. On construit des intervalles de confiance pour un seuil $\alpha = 5\%$.

1. Montrer que l'intervalle de confiance pour μ est de longueur $2 \times 1.96 \dfrac{\widehat{\sigma}}{\sqrt{n}}$
2. Montrer que l'intervalle de confiance pour σ est de longueur $2 \times 1.96 \dfrac{\widehat{\sigma}}{\sqrt{2n}}$ (en admettant qu'une loi $\chi^2(n-1)$ puisse s'approcher par une loi normale $\mathcal{N}(n-1, 2(n-1))$).

Exercice 3 – ***(Loi uniforme)***
On considère des observations $y_1, \cdots, y_n$ tirées suivant une loi $\mathcal{U}([0, \theta])$, avec $\theta > 0$.

1. Quel est l'estimateur du maximum de vraisemblance de θ, $\widehat{\theta}$?
2. Donnez la fonction de répartition de $\widehat{\theta}$.
3. À partir des quantiles de cette loi, proposez un intervalle de confiance pour θ.

Exercice 4 – ***(Compter les mots)***
Pour évaluer le nombre de mots d'un livre on tire 20 pages au hasard et on y compte le nombre de mots. On trouve, pour les 20 valeurs, une moyenne de 614 mots et un écart-type de 26 mots. Donner un intervalle de confiance à 95% pour le nombre total de mots du livre sachant qu'il a 158 pages (on admettra que l'approximation gaussienne est satisfaisante).

Exercice 5 – ***(Lois binomiales et binomiale négative)***
La loi binomiale négative est caractérisée par la fonction de probabilité

$$f(x; n, p) = \binom{x+n-1}{x} p^n (1-p)^x \quad \forall x = 0, 1, \cdots$$

et la loi binomiale

$$g(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x} \quad \forall x = 0, 1, \cdots, n$$

On dispose d'une seule observation x. Si n est supposé connu, donnez l'estimateur du maximum de vraisemblance pour p pour une loi binomiale négative et pour une loi binomiale.

Exercice 6 – ***(Poids à la naissance)***
On suppose que le poids d'un nouveau né est une variable normale d'écart-type égal à 0.5 kg. Le poids moyen des 49 enfants nés au mois de septembres 2021 dans l'hôpital de Laval a été de 3.6 kg.

1. Déterminer un intervalle de confiance à 95% pour le poids moyen d'un nouveau né dans cet hôpital.
2. Quel serait le niveau de confiance d'un intervalle de longueur (totale) 0.1 kg centré en 3.6 pour ce poids moyen ?

Exercice 7 – ***(Accidents de la route)***
Dans une ville on donne la répartition du nombre de jours sans accident, avec un accident, etc. parmi 50 jours d'observation au cours d'une même année :

Nombre d'accidents	0	1	2	3	4
Nombre de jours	21	18	7	3	1

On suppose que le nombre d'accidents par jour suit une loi de Poisson. Donner un intervalle de confiance de niveau 95% pour le nombre moyen d'accidents par jour (on utilisera une approximation asymptotique).

Exercice 8 – ***(Assurance qualité)***
Un stock comporte 10,000 pièces. Pour évaluer le nombre de pièces défectueuses dans le stock on tire au hasard 400 pièces dont on constate que 45 sont défectueuses. Donner un intervalle de confiance à 99% pour le nombre total de pièces défectueuses.

Exercice 9 – ***(Loi géométrique)***
Considérons un échantillon aléatoire $\{x_1, x_2, \cdots, X_n\}$ issu d'une loi géométrique de paramètre inconnu p dont la fonction de probabilité s'écrit

$$f(x;p) = p(1-p)^x \text{ pour } x = 0, 1, 2, \cdots$$

1. Montrez que $S = \sum_{i=1}^{n} X_i$ suit une loi binomiale négative de paramètres n et p.
2. On veut tester $H_0 :: p = 1/3$ contre $H_1 : p = 2/3$. Montrer que la région critique pour le test fondé sur le rapport de vraisemblance est de la forme $\sum_{i=1}^{n} X_i < k$
3. Donner la règle de décision pour $n = 4$ et $\alpha = 5\%$
4. Quelle est la puissance de ce test ?

Exercice 10 – ***(Naissance de prématurés)***
On sait que dans la population générale du nord de l'Italie le pourcentage de prématurés (naissance avant le 8ème mois) est de 4%. Dans une région du nord de l'Italie contaminée par une pollution chimique on a observé sur les dernières années 72 naissances prématurées sur 1 243 accouchements. Y-a-t-il lieu, selon la P-valeur constatée, de penser que la proportion de prématurés est plus élevée dans cette région que dans l'ensemble de la population du Nord du pays ? Donnez la fonction puissance du test de niveau 1%.

Exercice 11 – ***(Variation des dépenses de chauffage)***
L'an dernier, on a observé sur un échantillon de 29 appartements de 2 pièces situés en ville des dépenses de chauffage égales en moyenne à 325 \$, avec un écart-type égal à 26 \$. Cette année, pour un nouvel échantillon de 31 appartements de 2 pièces en ville on a trouvé des valeurs respectives de 338 \$ et 28 \$. L'hypothèse à laquelle on s'intéresse est qu'il n'y a pas eu d'augmentation des dépenses, en moyenne entre les deux années.

Exercice 12 – ***(Efficacité d'un vaccin)***
Un nouveau vaccin contre le paludisme est expérimenté auprès de la population d'une ville d'Afrique. On prend deux échantillons A et B de 200 personnes chacun. On injecte le vaccin aux individus de l'échantillon A et un placebo à ceux de l'échantillon B. Au bout d'un an on constate que 40 personnes de l'échantillon A ont des accès de palustres et 80 de l'échantillon B. Que dire de l'efficacité du vaccin ?

Exercice 13 – ***(Différence significative)***
Suite à des sondages, l'institut A donne 510 personnes favorables à telle mesure sur 980 personnes interrogées, l'institut B donne 505 favorables sur 1,030. La différence des estimations de la proportion de personnes favorables est-elle significative ?

Exercice 14 – ***(Gauchers et fumeurs)***
On dispose d'un échantillon de 1200 personnes, un tiers de gauchers, et deux tiers de droitiers. Parmi les gauchers, il y a 190 fumeurs, et 300 parmi les gauchers. La proportion des fumeurs est-elle la même parmi les gauchers et les droitiers ?

Exercice 15 – ***(Ampoules)***
En 1999, une étude sur la durée de vie des ampoules dans un amphi de cours a été lancée. Le 1er janvier 2000, toutes les ampoules ont été changées, et on a observé $n = 10$ durées de vie de ces n nouvelles ampoules, notées $X_1, \cdots, X_n$. On suppose que les variables X_i sont indépendantes, et de loi exponentielle, de moyenne θ (la durée est exprimée en mois).

1. Selon le constructeur, la durée de vie moyenne d'une ampoule est de 60 mois. Combien de temps en moyenne faut-il attendre avant que la premire ampoule ne meure ? Combien de temps en moyenne faut-il attendre avant que la dernire ampoule ne meurt ?
2. On retente l'expérience avec cette fois $n = 225$ ampoules. écrivez la vraisemblance du modle, $\mathcal{L}(\theta, \boldsymbol{X})$.
3. Que vaut l'estimateur du maximum de vraisemblance de θ.
4. Que vaut la variance asymptotique de l'estimateur du maximum de vraisemblance de θ.
5. A la fin de cette seconde expérience, on ne sait plus quel vendeur d'ampoule avait été sollicité. L'un prétendait vendre des ampoules d'une durée de vie moyenne de $\theta_0 = 60$ mois, et le second $\theta_1 = 66$ mois. On souhaite alors faire un test
$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1$$
On souhaite construire un test en utilisant la méthode du rapport de vraisemblance. Proposez une statistique de test, et la forme de la région de rejet.
6. Comme n est suffisamment grand, on admettra que le théorme central limite peut s'appliquer, et que $\overline{X}_n$ suit une loi normale. Donnez la loi normale approchée de $S_n = X_1 + \cdots + X_n$.
7. Donner la forme de la région de rejet pour une erreur de premire espce de niveau $\alpha = 5\%$. Pour rappel, $\Phi^{-1}(5\%) = -1.64$.
8. Calculez la puissance de ce test (donnez juste son expression en utilisant k).
9. Sur l'échantillon, on observe $\overline{X}_n = 62$. On souhaite tester
$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0 \text{ avec } \theta_0 = 60.$$
Expliquez graphiquement la construction du test de rapport de vraisemblance. Donnez la forme de la région de rejet.

Exercice 16 – ***(Loi Exponentielle)***
On dispose de n observations suivant une loi exponentielle $\mathcal{E}(\theta)$. On souhaite tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. écrire la vraisemblance et donnez la région de rejet du test de Neyman-Pearson de niveau $\alpha = 5\%$. Quelle est la puissance de ce test.
