Université de Rennes | Master 1
Outils en Proba stats | année 2023-2024

Feuille de Tds 6 : Estimateurs

$\boxed{1}$ Lois exponentielles et lois de Poisson. La durée écoulée entre l'arrivée de deux mails consécutifs dans la messagerie de Gérard suit une loi exponentielle de paramètre inconnu. On suppose que ces durées sont des réalisations de variables aléatoires indépendantes. L'observation est Y le nombre de mails arrivés dans la messagerie jusqu'à l'instant t.

1) Quelle est la loi de Y ?

2) Quel est le modèle statistique représentant cette expérience ?

$\boxed{2}$ D'une génération à la suivante, le poids moyen des cochons d'un élevage augmente d'un facteur multiplicatif inconnu, mais cette tendance est perturbée de façon additive par une variable aléatoire réelle gaussienne centrée, indépendante des générations précédentes et de variance inconnue indépendante de la génération. Quel est le modèle statistique associé à l'observation du poids moyen des cochons de l'élevage sur n générations ?

$\boxed{3}$ Soit le modèle statistique d'échantillonnage,

$$X_1, \ldots, X_n \text{ i.i.d. } \sim \mathcal{N}(m, \sigma^2),\ \theta = (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*.$$

Parmi les variables aléatoires suivante, lesquelles sont des statistiques ?
— la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$,
— $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$
— $n\hat{\sigma}^2/\sigma^2$,
— $T = \sqrt{n-1}\left(\bar{X} - m\right)/\hat{\sigma}$.

$\boxed{4}$ Pour chaque $\theta \in \mathbb{R}_+^*$, on note Q_θ la loi sur $\mathbb{R}_+$ de densité

$$f(x) = \theta e^{-x\theta} \mathbb{1}_{\mathbb{R}_+}(x).$$

On observe un n-échantillon de loi Q_θ.

1. Écrire le modèle statistique correspondant.

2. Est ce que le modèle est dominé ? Identifiable ?

3. Calculer l'espérance d'une v.a de loi Q_θ. Proposez un estimateur par la méthode des moments.

4. Soit $\hat{\theta} = \frac{n}{\sum_{i=1}^n X_i}$. Montrer que $\hat{\theta}$ est un estimateur consistant.

5. Montrer que $\hat{\theta}$ est un estimateur asymptotiquement normal c'est à dire qu'il existe $\sigma^2 > 0$ tel que

$$\sqrt{n}(\hat{\theta} - \theta) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0, \sigma^2)$$

et déterminer $\sigma^2 > 0$.

$\boxed{5}$ Rappel : Soit X une variable aléatoire de loi gaussienne d'espérance 0 et de variance 1. $\mathbb{E}(X^4) = 3$.

Pour chaque $\theta \in \mathbb{R}_+^*$, on note Q_θ la loi $\mathcal{N}(0, \theta)$ normal d'espérance 0 et de variance θ, où θ est un paramètre réel positif strictement positif. On observe un n-échantillon de loi Q_θ.

1. Écrire le modèle statistique correspondant.

2. Calculer le maximum de vraisemblance, noté $\hat{\theta}^{MV}$ de θ.

3. Calculer l'espérance mathématique et la variance de $\hat{\theta}^{MV}$. Que peut-on conclure ?

4. On suppose que le modèle vérifie les hypothèses afin de pouvoir appliquer les résultats sur l'information de Fisher. Calculer la quantité d'information de Fisher. En déduire que $\hat{\theta}^{MV}$ est efficace.

$\boxed{6}$ Soit le modèle d'échantillonage suivant :

$$X_1, ..., X_n \; i.i.d. \sim \mathcal{U}([\theta, \theta+1])^{\otimes n}, \;\; \theta \in [0, 1]$$

où $\mathcal{U}([\theta, \theta+1])$ est la loi uniforme sur $[\theta, \theta+1]$. Le paramètre d'intérêt est le paramètre de ce modèle.

1. Écrire le modèle sous la forme d'un triplet, espace d'observation, tribu, famille de probabilités.
2. Ce modèle est-il dominé ?
3. Le modèle est-il identifiable ?
4. Est ce que l'estimateur du maximum de vraisemblance de θ existe ?
5. Soit $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{2}$ un estimateur de θ. Est ce un estimateur biaisé ? Quelle est sa vitesse ?
6. Donner un intervalle de confiance asymptotique à 90 pourcent de θ.
7. Soit $\hat{\theta}_2 = \min_{i=1,\ldots,n} X_i$ un estimateur de θ. Quelle est la loi de $\hat{\theta}_2$?
8. Est ce $\hat{\theta}_2$ un estimateur biaisé ?
9. Quelle est sa vitesse de $\hat{\theta}_2$? Pour cela on pourra penser à calculer la fonction de répartition de la variable aléatoire $n(\hat{\theta}_2 - \theta)$ et étudier sa limite.
10. Question bonus : sur quel(s) critère(s) pourriez vous comparer ces deux estimateurs afin de savoir lequel est le plus pertinent ?

$\boxed{7}$ Soit $(X_1, ..., X_n)$ un n échantillon de variables aléatoires indépendantes de loi $\mathbb{P}_\theta$ pour $\theta \in]0;1[$, telle que pour tout entier k, on a $\mathbb{P}_\theta(X_1 = k) = \theta(1-\theta)^k$.

1. écrire le modèle statistique correspondant.

2. Calculer $\mathbb{E}(X_1)$ et déterminer un estimateur de θ par la méthode des moments.

3. Est-ce que le modèle est identifiable ?

4. Déterminer l'estimateur du maximum de vraisemblance.

5. On suppose que le modèle vérifie les hypothèses afin de pouvoir appliquer les résultats sur l'information de Fisher. Calculer la quantité d'information de Fisher.

Est-ce que $\hat{\theta} := \frac{1}{1+\overline{X}_n}$ est efficace ?

$\boxed{8}$ Des joueurs ont envie de créer un escape game à Liffré. Ils créent un sondage à Liffré dans le but de déterminer la proportion p d'individus ayant envie de faire un escape game.

On interroge 1000 personnes. Sur ces 1000 personnes, 352 affirment avoir envie de tenter l'expérience d'un escape game, et les autres n'en éprouvent pas l'envie. Sur la base de ces données, donner un intervalle de confiance à 90% pour p.

1) En utilisant l'inégalité de Bienaymé-Tchebychev.

2) En utilisant le théorème limite central.

$\boxed{9}$ Une chaîne de production doit garantir une qualité minimale de ses produits. En particulier, elle doit garantir que la proportion θ de produits défaillants reste inférieure à un taux fixé par le client. Un échantillon de n produits est prélevé et analysé. On appelle X_i la variable aléatoire définie par $X_i = 1$ si le $i$ème produit est defaillant, $X_i = 0$ sinon.On note $\hat{\theta}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ la proportion de produits défectueux dans l'échantillon.

1. Proposer un modèle statistique pour ce problème. Quelle est la loi de $n\hat{\theta}_n$?

2. Donner un intervalle de confiance asymptotique à 85 pourcent de θ.

$\boxed{10}$Lois exponentielles On considère le modèle statistique $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{E}(\lambda)^{\otimes n}{}_{\lambda>0})$. Dans la suite,

$(X_1, X_2, \ldots, X_n) \sim \mathcal{E}(\lambda)^{\otimes n}$ et $\alpha \in]0,1[$.

1. Montrer que la v.a. $\lambda \max_{1\leq i\leq n} X_i$ est pivotale, i.e. sa loi ne dépend pas de λ. En déduire un intervalle de confiance pour le paramètre λ au niveau de confiance $1-\alpha$.
2. Construire un intervalle de confiance pour λ au niveau $1-\alpha$ en utilisant la statistique $\bar{X}_n$.
3. Déterminer la vitesse et la loi limite de l'estimateur $\hat{\lambda} = 1/\bar{X}_n$. En déduire un intervalle de confiance asymptotique pour le paramètre λ au niveau $1-\alpha$.

$\boxed{11}$ Le coryza du chat est une maladie virale associée à un syndrome respiratoire qui touche principalement les chatons. Pour estimer la prévalence (c'est-à-dire la proportion de présence) dans un département, on prélève un échantillon de 145 chatons et 25 s'avèrent porteurs de cette maladie.

1. Calculer la fréquence observée de chatons porteurs du coryza dans cet échantillon.
2. Déterminer un intervalle de confiance au niveau de confiance asymptotique de 90% pour la proportion de chatons touchés par la maladie dans le département.
3. L'amplitude de l'intervalle de confiance asymptotique étant trop grande, on souhaite prélever un nouvel échantillon. Quel doit être le nombre de personne choisi pour l'échantillon si l'on veut que la taille de l'intervalle de confiance asymptotique soit inférieure ou égale à 0.01 ? On pourra donner la réponse sous la forme de produits et divisions de réels.

$\boxed{12}$ Soit $(X_1, ..., X_n)$ un n-échantillon de variables aléatoires indépendantes de loi $\mathbb{P}_\theta$ pour $\theta \in]0, 1[$ et $m \in \mathbb{N}^*$, telle que pour tout entier $k > 0$, on a $\mathbb{P}_\theta(X_1 = k) = \binom{k}{m}\theta^k(1-\theta)^{m-k}$. On suppose que m est connu.

1. Écrire le modèle statistique correspondant.
2. Calculer $\mathbb{E}_\theta(X_1)$ et déterminer un estimateur de θ par la méthode des moments.
3. Est-ce que le modèle est identifiable ?
4. Déterminer l'estimateur du maximum de vraisemblance.
5. Montrer que $Var_\theta(X_1) = m\theta(1-\theta)$. Vous pourrez utiliser cette valeur dans la suite même si vous n'avez pas réussi à le montrer.
6. Soit l'estimateur $\hat{\theta} := \frac{\overline{X_n}}{m}$. Quel est le MSE de $\hat{\theta}$?
7. Quelle est la vitesse de convergence notée c_n de $\hat{\theta}$? Que vaut $\sigma^2(\theta)$ dans l'expression $c_n(\hat{\theta} - \theta) \to \mathcal{N}(0, \sigma^2(\theta))$?
8. Donner un intervalle de confiance asymptotique à 85% de θ.
9. On suppose que le modèle vérifie les hypothèses afin de pouvoir appliquer les résultats sur l'information de Fisher. Calculer la quantité d'information de Fisher.
   Est-ce que $\hat{\theta}$ est efficace ?