Examen de cours de statistiques, ENPC, 08/09 ***Les documents et les calculatrices sont autorisés. Durée : 2h.***

(Attention : l'exercice 4 est différent selon le groupe des étudiants)

Exercice 1.

Un biochimiste étudie un type de moisissure qui attaque les cultures de blé. La toxine contenue dans cette moisissure est obtenue sous forme d'une solution organique. On mesure la quantité de substance par gramme de solution. Sur 9 extraits on a obtenu les mesures suivantes exprimées en milligrammes :

$$1. \quad 0.8 \quad 0.6 \quad 1.1 \quad 1.2 \quad 0.9 \quad 1.5 \quad 0.9 \quad 1.0$$

Déterminer un intervalle de confiance à 95% pour la moyenne de la quantité de substance toxique par gramme de solution (cette quantité de substance est supposée suivre une loi normale).

Exercice 2.

La hauteur maximale H de la crue annuelle d'un fleuve est observée car une crue supérieure à 6 mètres serait catastrophique. On a modélisé la distribution de la variable aléatoire H comme étant de Rayleigh, i.e la densité de H est donnée par

$$f(x) = \frac{x}{a} e^{-x^2/2a}, \quad x > 0.$$

où a est un paramètre inconnu. On suppose observer un n-échantillon de loi H que l'on note $H_1, \ldots, H_n$.

1. Montrer que $E(H_1^2) = 2a$.
2. Donner l'expression de la vraisemblance du modèle.
3. Calculer l'estimateur du maximum de vraisemblance $\hat{a}$ du paramètre a.
4. Montrer que l'estimateur du maximum de vraisemblance est sans biais.
5. Calculer, comme fonction du paramètre a, la probabilité qu'une crue dépasse 6 mètres. Pendant 8 ans on a observé les hauteurs de crue (en m)

$$2,5 \quad 2,9 \quad 1,8 \quad 0,9 \quad 1,7 \quad 2,1 \quad 2,2 \quad 2,8.$$

6. Une compagnie d'assurance estime qu'une catastrophe n'arrive en moyenne qu'au plus une fois tous les mille ans. Ceci peut-il être justifié par les observations ?

Exercice 3.

On se propose de comparer les réactions produites par deux vaccins B.C.G. désignés par A et B. Un groupe de 346 enfants a été divisé par tirage au sort en deux séries qui ont été vaccinées, l'une par A, l'autre par B. La réaction a été ensuite lue par une personne ignorant le vaccin utilisé. Les résultats figurent dans le tableau suivant :

Vaccin	Réaction légère	Réaction moyenne	Réaction forte	total
A	12	156	8	176
B	29	135	6	170
total	41	291	14	346

Pour chaque individu, on note par la variable $X \in \{A, B\}$ le vaccin qui lui a été appliqué, et par la variable $Y \in \{\text{Réaction lég.}, \dots, \}$ son état clinique.

1. Peut on vérifier que le tirage au sort est équitable (i.e. chaque individu à eu une chance sur deux d'être soumis à chacun des deux vaccins) ?
2. Tester si le vaccin appliqué a une incidence sur l'état clinique du patient (on pourra utiliser un test du chi2).

Exercice 4. (Cet exercice est uniquement pour les étudiants des groupes non IMI et pour les étudiants inscrits au premier semestre)

Soient T_1 et T_2 deux estimateurs sans biais d'un paramètre θ avec $var(T_1) = 2$, $var(T_2) = 3$ et $cov(T_1, T_2) = -1$.

1. A quelles conditions sur $(\lambda, \mu) \in \mathbb{R}^2$ l'estimateur $T = \lambda T_1 + \mu T_2$ est-il un estimateur sans biais ?
2. Parmi les estimateurs de la question précédente, lequel a un risque quadratique minimal ?

Exercice 4. (Cet exercice est uniquement pour les étudiants du groupe IMI)

On cherche à évaluer la proportion $p \in [0, 1]$ de personnes téléchargeant illégalement sur internet dans une population donnée. On choisit au hasard n individus dans cette population. La procédure de sondage est la suivante. On demande à chaque individu interrogé de tirer au préalable (sans être vu) une boule dans une urne, et de répondre ensuite par "OUI" ou "NON" à l'une des deux questions :

Question 1 : "*Est-ce que vous avez déjà telechargé illégalement ?*" si la boule tirée est blanche,

Question 2 : "*Est-ce que vous n'avez jamais téléchargé illégalement ?*" si la boule tirée n'est pas blanche.

(bien sur, seul l'individu interrogé a connaissance de la couleur de la boule, afin que personne d'autre ne sache à quelle question il a répondu)

La proportion ϕ de boules blanches dans l'urne est connue et différente de $1/2$. On note $X_1, \dots, X_n$ les variables aléatoires qui correspondent aux réponses des n personnes intérrogées (ce sont donc des variables aléatoires à valeur dans $\{\text{Oui}, \text{Non}\}$). Soit R le nombre de réponse "OUI" obtenues.

1. Quelle est la loi de X_1 ?
2. Quelle la loi de R ? Calculer $E(R)$?
3. En déduire un estimateur sans biais $\hat{p}$ de p.
4. Calculer le risque quadratique de l'estimateur $\hat{p}$. Commenter les cas $\phi = 0$, $\phi = 1$, $\phi \to 1/2$.