Série 5 - Adéquation statistique, distribution du χ^2, et niveau de signification

Thèmes: adéquation statistique, discrépance, distribution de probabilité du χ^2, niveau de signification. Table du χ^2. Table de la loi normale $\mathcal{N}(0,1)$.

Exercice 1

Un dé non biaisé a été lancé 100 fois avec les résultats suivants:

face k	1	2	3	4	5	6	total
fréquence observée f_k	16	15	14	20	22	13	100

Calculer la discrépance d et donner une estimation à l'aide de la table du χ^2 et donner le niveau de signifcation NS approximatif en utilisant la table. Est-ce que les données sont consistantes avec le modèle?

Indication, solution partielle: $d = 3.80$, NS ≈ 0.6, il n'y a donc pas d'évidence contre le modèle. Les données sont consistantes par rapport au modèle probabiliste de densité uniforme.

Exercice 2

Reprendre l'exercice sur la durée de vie des ampoules de la série 3. Considérer les classes mentionnées dans cet exercice et vérifier l'adéquation statistique avec le modèle exponentiel. On néglige le fait que les données ont été utilisées pour estimer le paramètre de la loi exponentielle. On demande de déterminer :

- le nombre de classes,
- le nombre de degrés de liberté,
- la discrépance observée,
- le niveau de signification observé en utilisant la distribution du χ^2,
- est-ce que les données récoltées sont-elles en désacord avec le modèle théorique ?

Exercice 3

Des pièces de plastiques sont testées pour leur résistance en leur administrant des coups de marteau répétitifs jusqu'à rupture. Le nombre de coups de marteau nécessaires avant cassure est supposée suivre une distribution géométrique de paramètre $p = \theta$ inconnu donnée par

$$f(x) = \theta^{x-1}(1-\theta) \qquad x = 1, 2, 3, \ldots$$

Les données récoltées sont :

nombre de coups de marteau k	1	2	3	≥ 4	total
nombre de pièces f_k	112	36	22	30	200

On introduit l'estimateur du paramètre θ comme

$$\hat{\theta} = \frac{f_2 + 2f_3 + 3f_4}{f_1 + 2f_2 + 3f_3 + 4f_4}$$

Est-ce que les données montrent une évidence contre le modèle de distribution géométrique ? Considérer que les données ont été utilisée pour estimer le paramètre θ.

Exercice 4

Pour combattre l'effet d'intoxication à l'alcool, une comparaison entre l'effet de la caféine et la benzédrine (amphétamine pharmaceutique, sulfate d'amphétamine, populaire entre 1930 et 1971) a été réalisée sous la forme d'un test clinique. Les deux substances sont administrées à 20 sujets, à deux reprises avec un ordre choisi au hasard pour la drogue administrée (benzédrine ou caféine), les deux essais bien espacés dans le temps. La benzédrine a permis un rétablissement plus prompt pour 14 cas, alors que la caféine a été plus efficace pour les 6 autres cas. Est-ce que ces résultats sont en adéquation statistique avec l'hypothèse que les deux drogues sont d'égale efficacité ? Quelle est la valeur du niveau de signification exact ? Quelle est la valeur du niveau de signification si on utilise la table de la loi normale ?

Indication, solution partielle: $X \triangleq$ le nombre de fois que la benzédrine est plus efficace que la caféine. Hypothèse $\mathscr{B}in(20, \frac{1}{2})$; NS $= P(D \geq 3.2) = P(X \leq 6) + P(X \geq 14) = 0.115$. $P(|Z| <?) =?$. Ces résultats ne contredisent pas de manière significative l'hypothèse d'égale efficacité entre les deux drogues.

Exercice 5

On demande dans cet exercice d'utiliser les tables, celle du χ^2 dans un premier temps, et celle de la loi $\mathscr{N}(0,1)$ dans un second temps.

Soit la discrépance $d = 1.32$ observée lors d'une expérience binomiale. On demande :

1. de déterminer le niveau de signification à l'aide de la table du χ^2;
2. d'utiliser l'approximation gaussienne pour déterminer le niveau de signification NS.

Indication, solution partielle: 1. NS $\approx$ 0.25; 2. NS $\approx$ 0.2502. La table de la loi normale indique $P(0 < Z \leq z)$ et on cherche $P(|Z| > \sqrt{d})$.
