# Tests du Khi-Deux

### Exercice 1 (L'expérience de Weldon)

Weldon a lancé 26306 fois un groupe de 12 dés. Il considérait comme succès l'apparition d'un 5 ou d'un 6 pour chacun des 12 dés. Une épreuve consiste en un jet des 12 dés. Dans chaque épreuve, on peut donc avoir $i$ succès, $1 \leqslant i \leqslant 12$. On note $N_i$ le nombre de fois ayant donné $i$ succès. Les résultats sont consignés dans le tableau ci-dessous :

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_i$ | 185 | 1149 | 3265 | 5475 | 6114 | 5194 | 3067 | 1331 | 403 | 105 | 14 | 4 | 0 |

1. En faisant l'hypothèse que les dés sont parfaits, calculer la probabilité d'obtenir $i$ succès lors d'une épreuve.
2. Déduire de ce qui précède la moyenne des épreuves qui ont fourni $i$ succès. Ces moyennes constituent les effectifs théoriques dans l'ajustement de la distribution réelle considérée par une loi binomiale.
3. En regroupant éventuellement certains effectifs de manière à ce qu'aucun effectif théorique ne soit inférieur à 5, compléter le tableau précédent par les effectifs théoriques $n_i$, par les différences $N_i - n_i$, et enfin par les quantités $\dfrac{(N_i - n_i)^2}{n_i}$. Calculer le nombre $\chi^2$ associé.
4. En utilisant le test du Khi-Deux, prouver que les dés sont mal construits, c'est-à-dire qu'ils ne satisfont pas à l'équiprobabilité des six faces.

---

### Exercice 2

Une population comprend quatre catégories d'individus indexées par 1,2,3,4 dans les proportions $p_1, p_2, p_3, p_4$. On effectue $n$ observations indépendantes ; soit $N_i$ le nombre d'individus de l'échantillon de la catégorie $i \in \{1, 2, 3, 4\}$.

1. Préciser la loi de $(N_1, N_2, N_3, N_4)$. Expliciter la fonction de vraisemblance de l'échantillon $(N_1, N_2, N_3, N_4)$ : $L(n_1, n_2, n_3, n_4) = \mathbb{P}(N_1 = n_1, N_2 = n_2, N_3 = n_3, N_4 = n_4)$.
2. Trois théories (non statistiques) s'affrontent à propos de la structure de la population, avec différentes valeurs des paramètres initiaux $p_1, p_2, p_3, p_4$ : des valeurs fixées (modèle I), puis des modèles paramétrés (modèles II et III) selon le tableau suivant :

| | | | | | |
|---|---|---|---|---|---|
| modèle I | : | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |
| modèle II | : | $p^3$ | $3p^2(1-p)$ | $3p(1-p)^2$ | $(1-p)^3$ |
| modèle III | : | $\frac{1}{8}+a$ | $\frac{3}{8}-a$ | $\frac{3}{8}-b$ | $\frac{1}{8}+b$ |

   (a) Écrire la fonction de vraisemblance $L(n_1, n_2, n_3, n_4; p)$ du modèle II.

   (b) Calculer $\dfrac{\partial}{\partial p} \ln L(n_1, n_2, n_3, n_4; p)$, puis déterminer l'estimateur du maximum de vraisemblance de $p$.

   (c) Écrire la fonction de vraisemblance $L(n_1, n_2, n_3, n_4; a, b)$ du modèle III.

   (d) Calculer $\dfrac{\partial}{\partial a} \ln L(n_1, n_2, n_3, n_4; a)$ et $\dfrac{\partial}{\partial b} \ln L(n_1, n_2, n_3, n_4; b)$, puis déterminer les estimateurs du maximum de vraisemblance de $a$ et $b$.

3. Les observations donnent pour $N_1, N_2, N_3, N_4$ les valeurs respectives suivantes : 121, 265, 316, 98. Qu'en est-il de l'ajustement de l'échantillon sur chacune de ces trois théories (au risque de 5 %).

---

### Exercice 3

On compte les arrivées de voitures à un péage sur l'autoroute pendant la durée d'une minute. On note $N_i$ le nombre d'intervalles de temps de 5 secondes où l'on a observé $i$ voitures, $0 \leqslant i \leqslant 11$. Le tableau suivant résume les 200 arrivées observées :

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $N_i$ | 1 | 15 | 30 | 46 | 38 | 30 | 16 | 13 | 5 | 3 | 2 | 1 |

On désigne par $X$ le nombre (aléatoire) d'arrivées au péage au cours d'une minute. On fait l'hypothèse que $X$ suit une loi de Poisson de paramètre $\theta$ inconnu.

1. Calculer la fonction de vraisemblance du paramètre $\theta$, $L(x_1, \ldots, x_n; \theta) = \prod_{k=1}^{n} \mathbb{P}(X = x_k)$, puis déterminer l'estimateur $T$ du maximum de vraisemblance pour $\theta$.
2. Étudier le biais et la convergence de $T$ puis donner une estimation de $\theta$.
3. Tester l'adéquation de la distribution empirique observée à la loi de Poisson estimée au seuil de 5 %.

---

**Exercice 4**

On a enregistré pendant $n = 100$ jours le nombre d'appels téléphoniques qui ont lieu de 9 h à 9 h 10 dans un certain central téléphonique. On a obtenu les résultats suivants :

| Nombre d'appels | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Nombre de jours | 2 | 14 | 23 | 24 | 18 | 9 | 6 | 2 | 1 | 0 | 1 |

(par exemple, il y a eu 18 des 100 jours pendant lesquels 4 appels ont eu lieu.) Soit $X$ la v.a. « nombre d'appels en un jour » et $(X_1, \ldots, X_n)$ un échantillon associé. On désigne par $m$ et $\sigma$ les moyenne et écart-type inconnus de $X$.

1. Construire le diagramme en bâtons correspondant. Calculer les valeurs observées de la moyenne $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et de la variance $V = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ de la v.a. $X$ .
2. On veut tester l'hypothèse suivante : « le nombre d'appels est distribué suivant une loi de Poisson ».
   (a) La relation entre la moyenne et l'écart-type mesurés est-elle en bon accord avec cette hypothèse ?
   (b) Quelle est la conclusion à laquelle conduit le test du $\chi^2$, avec un seuil de probabilité de 5 % ?
3. On suppose que la v.a. $\sqrt{n}\,(\overline{X} - m)/\sigma$ est normale conformément au théorème central limite. On refait alors la même statistique, portant également sur $n = 100$ jours, pour les appels téléphoniques qui ont lieu entre 10 h 30 et 10 h 40. On obtient pour ce nouvel échantillon noté $(Y_1, \ldots, Y_n)$ une moyenne $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Dans l'hypothèse où les appels sont distribués de la même façon vers 9 h et vers 10 h 30, trouver un nombre $M$ tel que l'on ait

   $$|\overline{Y} - \overline{X}| < M \text{ avec une probabilité de 95 \%.}$$

   Dans la détermination de $M$, on prendra pour valeur approchée de l'écart-type de la population la valeur déterminée expérimentalement sur le premier échantillon. (Suivant que les valeurs observées pour $\overline{X}$ et $\overline{Y}$ vérifieront ou non cette condition, on aura une présomption favorable ou non à l'égard de l'hypothèse précédente.)

---

**Exercice 5**

On désigne par $X$ la v.a. « nombre de pannes quotidiennes d'un appareil utilisé dans un laboratoire de recherche ». Les fréquences absolues de $X$ sur une période de 100 jours sont données par le tableau de correspondance suivant :

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $N_i$ | 45 | 30 | 16 | 6 | 3 |

1. Calculer la moyenne et l'écart-type de cette série statistique.
2. Construire le diagramme en bâtons des fréquences ce cette série. Quelle type de loi de probabilité semblerait ajuster convenablement la loi de $X$ ?
3. Justifier le choix précédent à l'aide d'un test du $\chi^2$.

---

**Exercice 6**

Une partie de la population d'une certaine institution est victime d'une maladie endémique (allergie hilbertienne) dont on distingue trois modalités de manifestation : hypocondrie [1], hyperesthésie [2], psychasténie [3] (*fluctuat nec mergitur*). Cette partie est constituée des affidés [4] prévaricateurs [5] de l'aréopage [6] Génie A, des

1. État d'anxiété permanente pathologique.
2. Exagération de la sensibilité, tendant à transformer les sensations ordinaires en sensations douloureuses.
3. État névrotique caractérisé par l'aboulie, l'asthénie, le doute, le scrupule et la méticulosité.
4. Personne à qui l'on se fie pour commettre une action répréhensible ; membre d'une société secrète, d'un complot.
5. Qui prévarique. Prévariquer : Manquer, par intérêt ou par mauvaise foi, aux devoirs de sa charge, de son mandat.
6. Assemblée de personnes choisies, particulièrement compétentes, savantes.

consorts[7] érémitiques[8] du cénacle[9] Génie B, des prosélytes[10] stipendiés[11] de la cabale[12] Génie C et des philistins[13] sybaritiques[14] du phalenstère[15] Génie D. Une enquête épidémiologique se propose d'étudier une éventuelle liaison entre le mode d'alimentation (nourritures gaussiennes) et la nature de l'endémie (*sustine et abstine*). Pour cela, on choisit au hasard dans chacune des quatre catégories précédentes un échantillon et pour chaque individu, on note l'intensité de la maladie. On constitue ainsi le tableau de contingence suivant :

| catégorie \ endémie | hypocondrie | hyperesthésie | psychasténie |
|---|---|---|---|
| Génie A | 105 | 89 | 51 |
| Génie B | 92 | 98 | 78 |
| Génie C | 97 | 83 | 105 |
| Génie D | 110 | 90 | 75 |

Au seuil de signification 0.01, faut-il écarter l'hypothèse d'indépendance des deux caractères : nature de l'endémie et mode d'alimentation ?

7. Personnes qui ont des intérêts communs, notamment dans une même procédure.
8. Relatif aux ermites.
9. Cercle restreint de personnes animées par des idées communes.
10. Personne gagnée à une cause, une doctrine, etc., et qui concourt à sa propagation.
11. Qui est payé pour accomplir une action ; corrompu.
12. Groupe des participants à une cabale. Cabale : Ensemble de menées secrètes dirigées contre quelqu'un, quelque chose.
13. Personne à l'esprit vulgaire, fermée aux lettres, aux arts, aux nouveautés.
14. Propre aux sybarites. Sybarites : Personne qui mène une vie facile et voluptueuse.
15. Vaste association de production au sein de laquelle les travailleurs vivent en communauté dans le système de Fourier.