Polytech Dijon | **21 novembre 2024**
3A MDD- S5 | **Documents autorisés**
Durée : 80 minutes | **Calculatrices autorisées**

OUTILS MATHÉMATIQUES I

Exercice 1 (10 points)

1. La fonction

$$L: \begin{array}{ccl} \mathbf{R}^3 & \to & \mathbf{R} \\ (\lambda_1, \lambda_2, \lambda_3) & \mapsto & e^{-200(\lambda_1+\lambda_2+\lambda_3)}\lambda_1^{12}\lambda_2^{14}\lambda_3^{16} \end{array}$$

a été définie dans l'énoncé. En tant que produit de fonctions différentiables sur $\mathbf{R}^3$, L est aussi différentiable sur $\mathbf{R}^3$. Pour tout $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ de $\mathbf{R}^3$ tel que $\lambda_1 \neq 0$

$$\begin{aligned} \frac{\partial L}{\partial \lambda_1}(\lambda) &= -200\exp(-200(\lambda_1+\lambda_2+\lambda_3))\lambda_1^{12}\lambda_2^{14}\lambda_3^{16} + \exp(-200(\lambda_1+\lambda_2+\lambda_3)) \cdot 12 \cdot \lambda_1^{11}\lambda_2^{14}\lambda_3^{16} \\ &= \exp(-200(\lambda_1+\lambda_2+\lambda_3))\lambda_1^{11}\lambda_2^{14}\lambda_3^{16} \cdot (-200\lambda_1 + 12) \\ &= \frac{L(\lambda)}{\lambda_1} \cdot (-200\lambda_1 + 12) \\ &= L(\lambda) \cdot \frac{12 - 200\lambda_1}{\lambda_1} \end{aligned}$$

2. La fonction de vraisemblance associées aux trois observations d'incidents dans les services A, B et C, les paramètres des trois lois de Poisson étant ici notés λ_1, λ_2 et λ_3, est

$$\frac{\lambda_1^{12}}{12!}e^{-200\lambda_1}\frac{\lambda_2^{12}}{14!}e^{-200\lambda_2}\frac{\lambda_3^{16}}{16!}e^{-200\lambda_3}$$

3. Pour calculer l'estimateur du maximum de vraisemblance, on cherche les paramètres $(\lambda_1, \lambda_2, \lambda_3)$ qui maximisent la vraisemblance ou, de façon équivalente, la fonction L.

$$\frac{\partial L}{\partial \lambda_1}(\lambda) = 0 \Rightarrow 12 - 200\lambda_1 = 0 \Rightarrow \hat{\lambda}_1^{MV} = \frac{12}{200} = \frac{3}{50}$$

On calculerait comme dans la question 1 la dérivée partielle de L par rapport à λ_B :

$$\frac{\partial L}{\partial \lambda_2}(\lambda) = L(\lambda) \cdot \frac{14 - 200\lambda_2}{\lambda_2}$$

donc

$$\frac{\partial L}{\partial \lambda_2}(\lambda) = 0 \Rightarrow 14 - 200\lambda_2 = 0 \Rightarrow \hat{\lambda}_2^{MV} = \frac{14}{200} = \frac{7}{100}$$

et enfin pour λ_C :

$$\frac{\partial L}{\partial \lambda_3}(\lambda) = L(\lambda) \cdot \frac{16 - 200\lambda_3}{\lambda_3}$$

donc

$$\frac{\partial L}{\partial \lambda_3}(\lambda) = 0 \Rightarrow 16 - 200\lambda_3 = 0 \Rightarrow \hat{\lambda}_3^{MV} = \frac{16}{200} = \frac{4}{50}$$

4.

$$\rho = \frac{\sup_{l \in \mathbf{R}} L(l,l,l)}{L(\lambda_1^{MV}, \lambda_2^{MV}, \lambda_3^{MV})}$$

On doit donc maximiser

$$F : l \mapsto e^{-200(l+l+l)}l^{12}l^{14}l^{16} = \exp(-600l)l^{42}$$

dont la dérivée vaut

$$F'(l) = -600\exp(-600l)l^{42} + 42\exp(-600l)l^{41} = \exp(-600l)l^{41}(42-600l)$$

Le maximum de F est atteint en l^* tel que

$$F'(l*) = 0 \Rightarrow 42 - 600l^* = 0 \Rightarrow l^* = \frac{42}{600} = \frac{7}{100}$$

Finalement,

$$\rho = \frac{\exp\left(-600\frac{7}{100}\right)\left(\frac{7}{100}\right)^{42}}{\exp\left(-200\frac{3}{50} - 200\frac{7}{100} - 200\frac{4}{50}\right)\left(\frac{3}{50}\right)^{12}\left(\frac{7}{100}\right)^{14}\left(\frac{4}{50}\right)^{16}} = \frac{\left(\frac{7}{100}\right)^{28}}{\left(\frac{3}{50}\right)^{12}\left(\frac{4}{50}\right)^{16}} = \frac{7^{28}}{6^{12}\cdot 8^{16}}$$

5. Testons

$$H_0 : \lambda_A = \lambda_B = \lambda_C \text{ contre } H_1 : \text{ les trois paramètres ne sont pas égaux}$$

Sous H_0, le rapport de vraisemblance $-2\ln\rho$ suit une loi χ^2 à autant de degrés de libertés que ceux spécifiés par l'hypothèse nulle. Sous H_1, il y a trois paramètres (λ_A, λ_B et λ_C). Sous H_0 il n'y en a plus qu'un (par exemple : λ_A) donc H_0 spécifie deux paramètres.

6. Ici, la réalisation de $-2\ln\rho$ vaut

$$-2\ln\frac{7^{28}}{6^{12}\cdot 8^{16}} = -2\left(28\ln(7) - 12\ln(6) - 16\ln(8)\right) = -56\ln(7) + 24\ln(6) + 32\ln(8) \approx 0.5734$$

qui correspond au quantile (environ) 75 % de la loi $\chi^2(2)$. La P-valeur est donc de l'ordre de 25 %, beaucoup trop élevée pour qu'on puisse rejeter l'hypothèse nulle.
En conclusion, les incidents constatés en 2024 s'expliquent très bien par une même probabilité d'incident pour les trois services et des fluctuations aléatoires. Ils ne permettent pas à eux seuls de mettre en cause la compétence des agents de tel ou tel service.

Exercice 2

Si une variable aléatoire X suit une loi exponentielle $\mathcal{E}(\lambda)$, alors

$$\mathbb{P}(X \leq 1) = \int_0^1 \lambda e^{-\lambda x}\mathrm{d}x = \left[-e^{-\lambda x}\right]_0^1 = 1 - e^{-\lambda}$$

Testons

$$H_0 : \text{les indications du vendeur sont correctes} \quad \text{contre} \quad H_1 : \text{les indications du vendeur sont correctes}$$

Sous H_0, alors $\lambda = \frac{1}{5}$ et le nombre N de disques en panne au bout d'un an suit une loi binomiale $\mathcal{B}(1000; 1-e^{-\frac{1}{5}})$. On cherche maintenant à savoir à quel quantile de la loi $\mathcal{B}(1000; 1-e^{-\frac{1}{5}})$ correspond l'observation $n = 400$. Cette valeur peut être donnée de manière approchée par une calculatrice (on trouve $\mathbb{P}(N \leq 399) = 1$, la différence avec 1 étant inférieure à la précision machine). Des amateurs de précision pourraient faire un calcul exact avec un calculateur formel.
Si on n'a pas de calculatrice sophistiquée, on peut utiliser une approximation normale. Ici $n = 1000$, $p = 1 - e^{-\frac{1}{5}} \approx 18.1\%$, donc $np \approx 181$, $np(1-p) \approx 148 >> 12$ et on peut très bien approcher la loi $\mathcal{B}(1000; 1-e^{-\frac{1}{5}})$ par la loi normale $\mathcal{N}(np; np(1-p))$

$$\mathbb{P}(N \geq 400) = \mathbb{P}\left(\frac{N-181}{\sqrt{148}} \geq \frac{399.5-181}{\sqrt{148}}\right) = \mathbb{P}\left(\frac{N-181}{\sqrt{148}}\right) \geq 17.91$$

Le valeur 17.01 n'est pas dans les tables usuelles de la fonction de répartition de la loi normale centrée réduite, donc $\mathbb{P}(N \geq 400) < 10^{-5}$.
En conclusion, les observations sont beaucoup trop surprenantes sous H_0 pour s'expliquer par le seul hasard. On rejette H_0 et on accepte H_1 : les indications du vendeur sont erronées.

Exercice 3

Face	1	2	3	4	5	6
Nombre de sorties	238	287	277	235	228	235

On note X_j la variable aléatoire donnant le résultat du lancer j. On supposera que les X_j sont indépendants et identiquement distribués. On notera p_k ($1 \leq k \leq 6$) la probabilité de sortie de la face k. Il n'y a en fait que cinq paramètres (pas 6) puisque $\sum_{k=1}^{6} p_k = 1$.
Pour $1 \leq j \leq 6$, notons S_j la variable aléatoire donnant le nombre de fois où la face j est sortie lors des 300 lancers, $S = (S_1, S_2, S_3, S_4, S_5, S_6)$ le vecteur aléatoire des résultats et $s = (s_1, s_2, \ldots, s_6)$ la réalisation observée de S. Le vecteur $(S_1, S_2, S_3, S_4, S_5, S_6)$ suit une loi multinomiale de paramètres 1500 et p=(p_1, p_2, p_3, p_4, p_5, p_6).
On veut tester

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6} \quad \text{contre} \quad H_1 : \exists k \in \{1,2,3,4,5,6\} | p_k \neq \frac{1}{6}$$

La vraisemblance s'écrit

$$L(s,p) = p_1^{s_1} \cdot p_2^{s_2} \cdot p_3^{s_3} \cdot p_4^{s_4} \cdot p_5^{s_5} \cdot p_6^{s_6}$$

donc ici

$$L(s,p) = p_1^{238} \cdot p_2^{287} \cdot p_3^{277} \cdot p_4^{235} \cdot p_5^{228} \cdot p_6^{235}.$$

Le maximum de vraisemblance est obtenu, d'après le calcul fait en cours, avec

$$\hat{p}_1 = \frac{238}{1500} \quad \hat{p}_2 = \frac{287}{1500} \quad \hat{p}_3 = \frac{277}{1500} \quad \hat{p}_4 = \frac{235}{1500} \quad \hat{p}_5 = \frac{228}{1500} \quad \hat{p}_6 = \frac{235}{1500}.$$

On calcule l'opposé du double du logarithme du rapport de vraisemblance :

$$\begin{aligned}
-2\ln\lambda &= -2\ln \frac{L\left(s, \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)\right)}{L\left(s, \left(\frac{238}{1500}, \frac{287}{1500}, \frac{277}{1500}, \frac{235}{1500}, \frac{228}{1500}, \frac{235}{1500}\right)\right)} \\
&= -2\ln \frac{\left(\frac{250}{1500}\right)^{238}\left(\frac{250}{1500}\right)^{287}\left(\frac{250}{1500}\right)^{277}\left(\frac{250}{1500}\right)^{235}\left(\frac{250}{1500}\right)^{228}\left(\frac{250}{1500}\right)^{235}}{\left(\frac{238}{1500}\right)^{238}\left(\frac{287}{1500}\right)^{287}\left(\frac{277}{1500}\right)^{277}\left(\frac{235}{1500}\right)^{235}\left(\frac{228}{1500}\right)^{228}\left(\frac{235}{1500}\right)^{235}} \\
&= -2\left(238\ln\left(\frac{250}{238}\right) + 287\ln\left(\frac{250}{287}\right) + 277\ln\left(\frac{250}{277}\right) + 235\ln\left(\frac{250}{235}\right)\right. \\
&\qquad \left. +228\ln\left(\frac{250}{228}\right) + 235\ln\left(\frac{250}{235}\right)\right) \\
&\approx 12.46
\end{aligned}$$

Justification de l'approximation normale On peut calculer $1500\frac{1}{6} = 250$, donc l'approximation normale est justifiée.
Sous H_0, les règles d'approximation étant vérifiées, $-2\ln\lambda$ suit (approximativement) une loi du χ^2 à cinq degrés de liberté, $\chi^2(5)$. D'après la table du formulaire, 12.46 correspond à une P-valeur comprise entre 2.5% et 5% (avec un tableur ou une calculatrice sophistiquée, on trouve une P-valeur d'environ 2.9%).
Conclusion : la P-valeur étant inférieure à 5%, on refuse l'hypothèse nulle. Notre conclusion est que les fluctuations du hasard ne suffisent pas à expliquer l'écart entre les observations et ce qui est attendu pour un dé non truqué : le dé est vraisemblablement truqué.

Exercice 4

Si on veut le traiter rigoureusement, l'exercice est plus difficile qu'on pourrait le crorie au premier abord. L'approche naïve donnée ci-dessous a rapporté l'intégralité des points.
Approche naïve
Comme il y a beaucoup d'observations entre 2000 et 2005 (on ne sait pas exactement combien, mais peut-être de l'ordre de $365 \times 6 > 2000$), on peut supposer que la valeur 50 μg/m^3 correspond réellement à l'espérance du maximum journalier entre 2000 et 2005.

Notons X_1, X_2, .. X_{365} les valeurs des maxima journaliers relevés en 2023 et $\bar{X}$ leur moyenne. D'après l'énoncé, on peut supposer que tous les X_i suivent une loi normale. Nous allons faire l'hypothèse supplémantaire forte que tous ces relevés suivent la *même* loi $\mathcal{N}(\mu, \sigma^2)$ et sont *indépendants entre eux*. Donc $\bar{X}$ suit une loi $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
On va tester

$$H_0 : \mu \geq 50 \text{ contre } H_1 : \mu < 50$$

(on aurait aussi pu tester l'égalité avec un test bilatéral mais comme la question porte sur une baisse, on préfére tester spécifiquement dans le sens qui nous intéresse)
<u>Sous H_0</u>, la statistique

$$T = \frac{\bar{X} - 50}{s/\sqrt{365}}$$

suit une loi de Student à 364 degrés de liberté, donc en pratique une loi normale centrée réduite (sauf à disposer de calculatrices sophistiquées ou d'ordinateurs). On calcule $t \approx -5.6$ et le formulaire donne une P-valeur de pratiquement 0 (on trouve $9.6 \cdot 10^{-9}$ avec un tableur).
<u>Conclusion</u> : la P-valeur est trés largement inférieure à 5 %. Avec une quasi certitude on peut dire que les fluctuations du hasard ne suffisent pas, sous l'hypothése $\mu = 50$, à expliquer l'écart entre les relevés et l'attendu sous H_0. On rejette donc H_0 et on déduit que la moyenne des maxima journalier a baissé.

Approche plus rigoureuse
On ne connaît pas exactement la loi des relevés (dont on ne connaît pas le nombre) pour la période 2000-2005. On peut suivre l'énoncé et suppose que les relevés $2000-2005$ suivent une loi normale $\mathcal{N}(\mu_1, \sigma_1^2)$. Si on considère que ces relevés sont indépendants (hypothèse raisonnable) et au nombre de $6 \cot 365 + 2 =$ (un relevé par jour, 2000 et 2004 sont des années bissextiles), alors la moyenne $\bar{X}_1$ de ces relevés suit une loi $\mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{2192}\right)$. Pour tester

$$H_0 : \mu \geq \mu_1 \qquad \text{contre} \qquad H_1 : \mu < \mu_1,$$

on peut dire que sous H_0, $\bar{X} - \bar{X}_1$ suit une loi normale $\mathcal{N}\left(\mu - \mu_1, \frac{\sigma_1^2}{2192} + \frac{17^2}{365}\right)$. On ne peut pas aller plus loin si on ne connaît pas σ_1. Si on suppose que l'écart-type est encore de 17 μg/m^3 (ceci est une hypothèse très forte), alors on connaît la loi de $\bar{X} - \bar{X}_1$ (c'est une loi normale $\mathcal{N}(\mu - \mu_1, 0.92)$ et on peut calculer la P-valeur (environ 10^{-7}), et on aboutit à la même conclusion que précédemment en rejetant H_0 : le niveau de pollutiton a diminué.
Sans hypothèse sur σ_1, la question devient très difficile. À titre d'exercice, on peut calculer quelle devrait-être la valeur de σ_1 pour qu'on ne puisse pas conclure avec un risque de 5%. Si la valeur observée $\bar{x} - \bar{x}_1$

Exercice 5 (5 points)
Il s'agit d'estimer le paramètre d'une loi binomiale. Notons X_1,..X_{577} les résultats (0=pile, 1=face) des lancers. Si on suppose que tous les lancers sont indépendants et ont tous la même probabilité p de donner « face », alors le nombre de face S total suit une loi binomiale $\mathcal{B}(577, p)$. L'estimteur du maximum de vraisemblance pour p est $\hat{p} = \frac{203}{577}$.
Remarquons que $n \cdot \hat{p} \cdot (1 - \hat{p}) \approx 131 >> 12$ et l'approximation normale est parfaitement justifée. On peut utiliser directement les formules du cours

$$I_{0.95}(p) = \left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{577}}, \hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{577}}\right] =]0.31, 0.39[$$

Si on dispose d'une calculatrice sophistiquée, on peut utiliser une méthodes quantiles et trouver un intervalle identique.
On constate que 0.5 n'est pas dans l'intervalle de confiance à 95% pour p : on peut rejeter l'hypothèse $p = 0.5$ au niveau de risque 5%.

Exercice 6 (5 points)

Il s'agit de tester l'égalité des paramètres entre deux lois multinomiales. La seule méthode à notre disposition pour l'instant est un test du rapport de vraisemblance généralisé, valable pour les grands échantillons ce qui est le cas ici.
On dispose de deux dés. On suppose que les sorties suivent des lois multinomiales de paramètres p_1^A, $p_2^A..p_6^A$ pour le dé 1 et p_1^B, $p_2^B..p_6^B$ pour le dé 2.
On veut tester

$$H_0 : p_j^A = p_j^B \quad \forall j \in \{1..6\} \qquad \text{contre} \qquad H_1 : \exists j \in \{1..6\} | p_j^A \neq p_j^B$$

Sous H_0, l'estimateur du maximum de vraisemblance donne

$$\begin{aligned}
p_1^A = p_1^B &= \frac{238+236}{3000} = 0.158 \\
p_2^A = p_2^B &= \frac{287+283}{3000} = 0.19 \\
p_3^A = p_3^B &= \frac{277+279}{3000} = 0.185 \\
p_4^A = p_4^B &= \frac{235+239}{3000} = 0.158 \\
p_5^A = p_5^B &= \frac{228+225}{3000} = 0.151 \\
p_6^A = p_6^B &= \frac{235+238}{3000} = 0.158
\end{aligned}$$

L'estimateur du maximum vraisemblance sans faire l'hypothèse H_0 donne

$$\begin{aligned}
p_1^A &= \frac{238}{3000} = 0.159 & p_1^B &= \frac{236}{3000} = 0.157 \\
p_2^A &= \frac{287}{3000} = 0.191 & p_2^B &= \frac{283}{3000} = 0.189 \\
p_3^A &= \frac{277}{3000} = 0.185 & p_3^B &= \frac{279}{3000} = 0.186 \\
p_4^A &= \frac{235}{3000} = 0.157 & p_4^B &= \frac{239}{3000} = 0.159 \\
p_5^A &= \frac{228}{3000} = 0.152 & p_5^B &= \frac{225}{3000} = 0.150 \\
p_6^A &= \frac{235}{3000} = 0.157 & p_6^B &= \frac{238}{3000} = 0.159
\end{aligned}$$

Le calcul de $\ln \rho$, où ρ désigne le rapport de vraisemblance, donne

$$\ln \rho = 474 \ln(0.158) + 570 \ln(0.19) + 556 \ln(0.185) + 474 \ln(0.158) + 453 \ln(0.151) + 473 \ln(0.158)$$

Finalement, $-2 \ln \rho \approx$
On lance 1500 fois deux dés à 6 faces (numérotées de « 1 » à « 6 »), on compte pour chaque dé le nombre de fois où chaque face est obtenue. Les résultats sont consignés dans le tableau ci-dessous. (Lecture du tableau : le « 1 » a été obtenu 238 fois avec le dé 1, et 236 fois avec le dé 2.)

Face	1	2	3	4	5	6
Dé 1	238	287	277	235	228	235
Dé 2	236	283	279	239	225	238

Peut-on raisonnablement penser que les deux dés sont identiques ?