Final SQ20

Jeudi 24 juin 2021, 8h00-9h45

Le barème indicatif est 8+12 points pour les exercices 1 et 2. Un exercice additionnel vous est proposé en fin de sujet.

Exercice 1 *Qui fera la vaisselle ?.* Alice, Bob et Charlie partagent une colocation et jouent tous les soirs à "Qui fera la vaisselle ?".

Charlie sort un dé de sa poche et leur propose le jeu suivant : l'un d'entre eux lance le dé, si le dé affiche 1 où 4, Alice fait la vaisselle, si le dé affiche 2 ou 5, Bob fait la vaisselle, dans les autres cas ce sera lui qui fera la vaisselle. Soit X la variable aléatoire qui prend la valeur 0 si Alice fait la vaisselle, 1 si Bob fait la vaisselle et 2 si Charlie fait la vaisselle. Charlie n'a pas précisé à ses colocataires que le dé est pipé et que la loi de X est donnée par la distribution de probabilité

$$\mathbb{P}(X=0)=\mathbb{P}(X=1)=\frac{\theta}{2} \tag{1}$$

avec $\theta \in]0,1[$.

1. Calculer l'espérance et la variance de X. [1]
2. Pour quelles valeurs de θ le jeu est-il favorable à Charlie ?
3. Après n jours passés à jouer la vaisselle, Alice et Bob s'étonnent de faire plus souvent la vaisselle que Charlie et décide de procéder à une étude statistique.
   Soient $X_1,\dots,X_n$ un n-échantillon de distribution de probabilité donnée par (1). On note N_j, pour $j \in \{0,1,2\}$, les variables aléatoires qui représentent le nombre de fois où Alice, Bob et Charlie respectivement auront fait la vaisselle au bout de n jours.
   (a) Soit la statistique
   $$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$
   Quelle est son espérance et sa variance ?
   (b) Proposer un estimateur sans biais T_n de θ basé sur la statistique $\bar{X}_n$. [2] Quel est son espérance et sa variance ?
   (c) Montrer que $\bar{X}_n = (N_1+2N_2)/n$ puis exprimer T_n en fonction des N_j, $j \in \{0,1,2\}$.
   (d) Calculer en fonction des N_j, $j \in \{0,1,2\}$, l'estimateur $\hat{\theta}_n$ de θ basé sur la méthode du maximum de vraisemblance. [3] Vous pourrez admettre que si n_0, n_1 et n_2 représentent le nombre de fois où Alice, Bob et Charlie ont fait la vaiselle ($n_0+n_1+n_2=n$) alors la vraisemblance de ces observations est (loi multinomiale)
   $$L(n_0,n_1,n_2;\theta) = \text{ constante} \times \left(\frac{\theta}{2}\right)^{n_0+n_1}(1-\theta)^{n_2}.$$
   (e) En remarquant que $N_0+N_1 \sim \mathscr{B}(n;\theta)$, donner l'espérance et la variance de $\hat{\theta}_n$.
   (f) Montrer que $\mathrm{Var}(T_n) - \mathrm{Var}\left(\hat{\theta}\right) > 0$ pour $0<\theta<1$. Conclusion ?
   (g) Au bout de $n=100$ jours, les valeurs observées par Alice et Bob sont $n_1=27$ et $n_2=23$. Quelles sont les estimations ponctuelles de θ données par T_n et $\hat{\theta}_n$?

1. Vous devriez trouver $2-\frac{3}{2}\theta$ et $\frac{\theta(10-9\theta)}{4}$.
2. T_n est de la forme $a\bar{X}_n+b$ avec a,b deux constantes qui ne dépendent pas de θ
3. Vous devriez trouver $\frac{N_0+N_1}{n}$.

Exercice 2 *Consommation d'eau et canicule.* La consommation journalière en eau d'une agglomération au cours du mois de juillet est une variable aléatoire X dont la densité $f_X(x;\theta)$ a la forme :

$$f_X(x;\theta) = cx(\theta - x)1_{[0,\theta]}(x), \qquad x \in \mathbb{R}, \tag{2}$$

où c et θ sont des constantes strictement positives et $1_{[0,\theta]}(x) = 1$ si $x \in [0,\theta]$ et 0 sinon.

On vous rappelle que pour tout $n \in \mathbb{N}$

$$\int_0^\theta x^n(\theta - x)dx = \frac{\theta^{n+2}}{(n+1)(n+2)}. \tag{3}$$

1. **Partie 1 : Étude d'une densité continue**
   (a) Exprimer la constante c en fonction de θ.
   (b) Calculer $\mathbb{E}[X]$ et $\mathbb{E}\left[X^2\right]$. En déduire la variance de la variable aléatoire X.
   (c) Montrer que la fonction de répartition $F_X(x;\theta)$ de la variable aléatoire X est la suivante

$$F_X(x;\theta) = \frac{6}{\theta^3}x^2\left(\frac{\theta}{2} - \frac{x}{3}\right) \tag{4}$$

   On pourra se contenter de dériver F_X
   Proposer une interprétation physique de la constante θ.
   (d) En notant X_i la consommation du i-ème jour et en supposant que les X_i sont indépendantes et de même loi que X, exprimer à l'aide de F_X et de n la fonction de répartition de la variable aléatoire $M_n = \max_{1\le i\le 31} Xi$.
   (e) En fait la ville est alimentée en eau par un aqueduc qui peut fournir au maximum une quantité journalière d'eau $x_0 = 0.9\times\theta$ et par un réservoir de sécurité dans lequel elle peut puiser en cas de trop forte demande. Calculer numériquement la probabilité qu'au cours des 31 jours du mois de juillet, on ne fasse jamais usage du réservoir de sécurité (le résultat ne dépend pas de θ).

2. **Partie 2 : Statistiques**
   La valeur de θ a été déterminée en 2010 par une étude statistique et fixée à la valeur de 480 (en milliers de litre). Ce nombre semble obsolète et la régie des eaux de la ville cherche à réaliser une nouvelle estimation de θ en utilisant comme observations n consommations journalières de l'année 2020 notées $(x_1, x_2, \ldots, x_n)$.
   Soit $(X_1, \ldots, X_n)$ une suite de variable aléatoire iid de densité de probabilité $f_X(.;\theta)$.
   (a) On considère la statistique

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   Quelle est son espérance et quelle est sa variance ?
   (b) Déduire de cette statistique un estimateur sans biais T_n de θ dont vous donnerez l'espérance et la variance.
   (c) Cet estimateur est-il convergeant ?
   (d) Si on a observé en juillet 2020 une consommation totale de 9300 milliers de litre, quelle estimation ponctuelle de θ proposeriez-vous ?
   (e) Proposer un intervalle de confiance au niveau 0,95 pour θ en utilisant le TCL et en remplaçant $\mathrm{Var}(T_n)$ par son estimation. Préciser bien les hypothèses que vous faîtes.

Exercice 3 *Lois bivariées : les mauvais et les bons conducteurs (Exercice en option).* Pour les assurances, le nombre N de sinistres causés en une année par un assuré donné est souvent modélisé par une loi de Poisson de paramètre α. On a alors $\mathbb{E}[N] = \alpha$, ce qui permet d'interpréter le paramètre α comme un nombre moyen de sinistres. Ce modèle ne colle pas bien à la réalité lorsque l'on étudie l'ensemble des assurés correspondant à une police donnée (par exemple les conducteurs de véhicule d'un certain type dans une certaine zone géographique). Pour prendre en compte l'hétérogénéité des assurés, on remplace alors α par $R\alpha$ où R est une variable aléatoire positive d'espérance 1, appelée niveau de risque relatif. Le modèle « bons risques – mauvais risques » en assurance automobile est le plus simple de ce type. Il revient à partager l'ensemble d'assurés étudié en « bons conducteurs » pour lesquels le nombre de sinistres suit la loi de Poisson de paramètre $r_1\alpha$ (avec $r_1 < 1$) et en « mauvais conducteurs » pour lesquels le nombre de sinistres suit la loi de Poisson de paramètre $r_2\alpha$ avec $r_2 > 1$. Le problème est qu'il est difficile de décider au vu de l'historique des accidents qui est bon ou mauvais conducteur. Parmi les assurés n'ayant pas causé de sinistre lors des années précédentes, il peut très bien y avoir des mauvais conducteurs « chanceux » et parmi les assurés ayant causé au moins un sinistre, il peut tout aussi bien y avoir des bons conducteurs « malchanceux ». On considère donc que le niveau de risque relatif d'un assuré choisi au hasard est une variable aléatoire R vérifiant :

$$\mathbb{P}(R = r_1) = p_1 > 0,\ \mathbb{P}(R = r_2) = p_2 > 0, \qquad p_1 + p_2 = 1, \qquad \mathbb{E}[R] = 1. \tag{5}$$

Les paramètres p_i, r_i peuvent être estimés par l'observation approfondie des comportements au volant d'un échantillon d'assurés. Dans toute la suite, nous les supposerons connus. Le seul paramètre inconnu de notre modèle sera donc α. On s'intéresse au nombre N de sinistres causés par un assuré choisi au hasard (on ne sait pas s'il est bon ou mauvais conducteur).

1. Exprimer la condition $\mathbb{E}[R] = 1$ à l'aide de r_1, r_2 et p_1.
2. Quelle est la loi de N conditionnelle à l'événement $\{R = r_1\}$? Même question en conditionnant par l'événement $\{R = r_2\}$.
3. En déduire que la loi de N est donnée par

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(N = k) = p_1 \frac{e^{-r_1\alpha}(r_1\alpha)^k}{k!} + p_2 \frac{e^{-r_2\alpha}(r_2\alpha)^k}{k!}. \tag{6}$$

   On dit que N suit la loi de *Poisson mélange* de moyenne α et de niveau de risque relatif R donné par (5).
4. Calculer l'espérance de N ainsi que sa variance
5. On considère désormais le modèle statistique $(\Omega, \mathscr{F}, (\mathbb{P}_\alpha)_{\alpha \in]0,+\infty[})$ et un n-échantillon

$$N_1, \ldots, N_n$$

   associé. Autrement dit, pour toute valeur de α, les variables aléatoires N_i sont indépendantes et de même loi donnée par (6). On note $\overline{N}_n$ la moyenne empirique de cet échantillon. Justifiez brièvement [4] les deux convergences suivantes :

$$\forall \alpha > 0, \quad \overline{N}_n \xrightarrow[n\to+\infty]{proba} \alpha; \tag{7}$$

$$\forall \alpha > 0, \quad \sqrt{\frac{n}{\mathrm{Var}(N_1)}}\left(\overline{N}_n - \alpha\right) \xrightarrow[n\to+\infty]{loi} Z, \qquad Z \sim \mathcal{N}(0,1). \tag{8}$$

4. On se contentera de rappeler les deux théorèmes fondamentaux des probabilités vus en cours

6. On admettra qu'on a aussi la convergence en loi suivante où l'on a posé $\rho^2 = \mathrm{Var}(R)$ (la loi de R est connue et ne dépend pas de α, voir (5)) :

$$\forall \alpha > 0, \quad \sqrt{\frac{n}{\frac{1}{n} + \overline{N}_n(1+\rho^2\overline{N}_n)}}\left(\overline{N}_n - \alpha\right) \xrightarrow[n\to+\infty]{loi} Z, \qquad Z \sim \mathcal{N}(0,1). \tag{9}$$

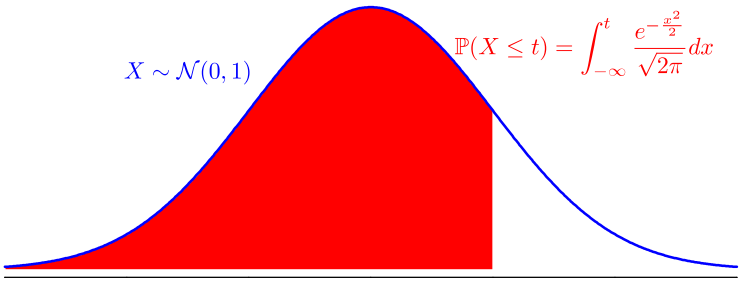
   En déduire un intervalle de confiance I au niveau 95% pour α. Pouvait-on utiliser (8) pour obtenir un tel intervalle ?

7. Le tableau ci-dessous donne un échantillon observé de taille 100, la dernière colonne contient la somme des lignes.

0	1	0	2	1	0	0	2	0	0	0	0	1	0	0	1	0	0	2	0	10
0	1	1	0	0	0	2	0	0	1	0	1	0	1	1	0	0	0	2	1	11
1	0	0	2	0	0	1	1	1	0	1	0	0	1	0	0	0	1	0	0	9
0	3	1	0	0	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	10
1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	2	0	3	10

   On « sait » que $p_1 = 0,25$, $r_1 = 0,7$ et $r_2 = 1,1$. Donnez l'estimation ponctuelle de a et l'intervalle de confiance I calculé à partir de cet échantillon.

Annexes



Loi normale centrée réduite

t	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table pour les grandes valeurs de t :

t	3	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
F(t)	0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,99984	0,99989	0,99993	0,99995

t	4	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9
F(t)	0,99997	0,99998	0,99999	0,99999	0,99999	1,00000	1,00000	1,00000	1,00000	1,00000