Contrôle : Statistique inférentielle

Patrick Tardivel, Université de Bourgogne

19/05/2023

Exercice 1. *(2,5 points) On teste un nouveau traitement contre le cancer du sang. Pour un individu i ayant suivi le traitement, on note $X_i = 1$ lorsque l'individu guéri et $X_i = 0$ sinon. On souhaite déterminer la probabilité $p \in]0,1[$ inconnue de guérir après traitement.*

1. *Rappeler la démarche mathématiques permettant de construire un intervalle de confiance asymptotique pour p au niveau $1-\alpha$.*
2. *Pour 720 individus traités on observe 305 guérisons. Calculer l'intervalle de confiance expérimental pour p de niveau asymptotique 0,90.*

Exercice 2. *(3,5 points) On souhaite savoir si le poids moyen des hyènes femelles de la réserve nationale du Masai Maras au Kenya est différent du poids moyen des hyènes femelles du parc national du Serengeti en Tanzanie.*
Données : *On a pesé 187 hyènes au Kenya dont les poids sont $x_1, x_2, \dots, x_{187}$. On a obtenu les résultats suivants :*

$$\sum_{i=1}^{187} x_i = 12171,33 \text{ et } \sum_{i=1}^{187} x_i^2 = 800253,8.$$

On a pesé 127 hyènes en Tanzanie dont les poids sont $y_1, y_2, \dots, y_{127}$. On a obtenu les résultats suivants :

$$\sum_{i=1}^{127} y_i = 7852,12 \text{ et } \sum_{i=1}^{127} y_i^2 = 493556,9.$$

1. *Faire un test statistique au niveau 5% pour conclure en rédigeant soigneusement chaque étape.*
2. *Calculer la p-valeur.*

Exercice 3. *(5 points) On considère une méthode pseudo-aléatoire pour générer un échantillon d'une loi uniforme sur les chiffres $\{0, 1, \dots, 9\}$ basée sur la suite de Fibonacci : $u_{n+2} = u_{n+1} + u_n \bmod 10$. En prenant pour valeurs initiales $u_1 = 3$ et $u_2 = 5$, les 200 premiers chiffres de cette suite fournissent le tableau suivant*

Chiffre	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Nombre d'apparitions	*13*	*27*	*12*	*27*	*15*	*27*	*13*	*28*	*13*	*25*

1. *Faire un test statistique d'adéquation à une loi uniforme sur $\{0, 1, \dots, 9\}$ au niveau 5% pour conclure en rédigeant soigneusement chaque étape.*
2. *Calculer la p-valeur (vous pouvez utiliser la table donnée à la page 4 pour calculer la p-valeur).*
3. *En observant attentivement le tableau, il semblerait que la fréquence d'obtention des chiffres impairs soit le double de la fréquence d'obtention des chiffres pairs. Est-ce qu'un test statistique contredit cette affirmation ? (le calcul de la p-valeur n'est pas demandé).*

Exercice 4. *(9 points) On souhaite comparer deux méthodes stochastiques d'approximation du nombre d'or $\Phi = (1+\sqrt{5})/2$. On note $\mathbf{1}_A$ la variable aléatoire indicatrice de l'événement A ($\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon).*

1. *Soit U une variable aléatoire de loi uniforme sur $[0,1]$. Montrer que $\mathbb{P}(U^2 + U - 1 \leq 0) = \Phi - 1$*
2. *Soit $U_1, \dots, U_n$ une suite de variable aléatoire indépendantes de même loi uniforme sur $[0,1]$. D'après la question 1, il est naturel d'estimer Φ via la statistique suivante*

$$T_1 = 1 + \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{U_i^2 + U_i - 1 \leq 0}.$$

   Calculer le biais, la variance et l'erreur quadratique moyenne pour Φ de T_1.

3. *Soit U une variable aléatoire de loi uniforme sur $[0,1]$, calculer $\mathbb{P}(U^2 \leq 5/9)$.*
4. *Soit $U_1, \dots, U_n$ une suite de variable aléatoire indépendantes de même loi uniforme sur $[0,1]$. D'après la question 3, il est naturel d'estimer Φ via la statistique suivante*

$$T_2 = 0,5 + \frac{3}{2n} \sum_{i=1}^{n} \mathbf{1}_{U_i^2 \leq 5/9}.$$

Calculer le biais, la variance et l'erreur quadratique moyenne pour Φ de T_2.

5. *Quel est le meilleur estimateur de Φ entre T_1 et T_2 ?*
6. *Quelle est la loi asymptotique de la statistique*

$$\frac{T_1 - \Phi}{\sqrt{\frac{(T_1-1)(2-T_1)}{n}}}$$

lorsque n tend vers $+\infty$? Vous Justifierez votre réponse.

7. *On considère le programme sur R suivant :*

```
A=runif(10000)
B=1+mean(A^2+A-1<0)
I0=B-1.645*sqrt((B-1)*(2-B))/100
I1=B+1.645*sqrt((B-1)*(2-B))/100
```

Que représentent les variables B, I0, I1 ?

8. *En vous inspirant des questions 6 et 7, écrire un programme qui, à partir d'un échantillon expérimental d'une loi uniforme sur $[0,1]$ de taille $n = 10000$, renvois un intervalle de confiance expérimental pour Φ basé sur l'estimateur T_2 de niveau asymptotique $0,95$. Vous justifierez votre démarche.*

Formulaire

— Soit $X_1, \dots, X_n$ un $n-$échantillon d'une loi P inconnue à valeur dans un ensemble fini D et P^0 une loi de probabilité sur D. Pour $i \in D$ on pose $N_i = \sum_{j=1}^{n} \mathbf{1}(X_i = i)$. Sous l'hypothèse nulle, lorsque $P^0 = P$, on a la convergence en loi suivante :

$$Y = \sum_{i \in D} \frac{(N_i - nP^0(i))^2}{nP^0(i)} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi(\text{card}(D) - 1 \text{ ddl})$$

— Soit $(X_1, \dots, X_n)$ un $n-$échantillon d'une loi de moyenne $\mu_1 \in \mathbb{R}$ et de variance σ_1^2 indépendant de $(Y_1, \dots, Y_n)$ un $n-$échantillon d'une loi de moyenne $\mu_2 \in \mathbb{R}$ et de variance σ_2^2 alors, sous l'hypothèse nulle lorsque $\mu_1 = \mu_2$, on a la convergence en loi suivante :

$$Z := \frac{\overline{X} - \overline{Y}}{\sqrt{S^2_{corr}(X)/n_1 + S^2_{corr}(Y)/n_2}} \xrightarrow[\min\{n_1, n_2\} \to +\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

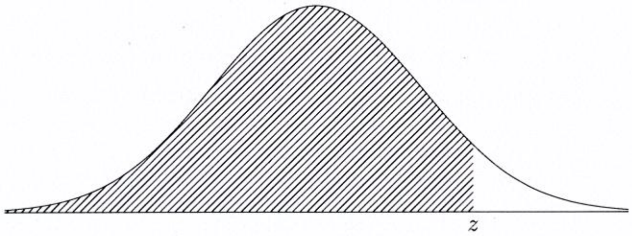
— Soit $(X_1, \dots, X_n)$ un $n-$échantillon de loi $\mathcal{B}(p)$ alors :

$$\frac{\overline{X} - p}{\sqrt{\overline{X}(1 - \overline{X})/n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Table de la loi normale centrée réduite

Soit Z une variable aléatoire de loi $\mathcal{N}(0,1)$. Les valeurs de la fonction de répartition $z \in \mathbb{R} \mapsto \mathbb{P}(Z \leq z)$ sont données dans le tableau suivant :

densité d'une loi normale $\mathcal{N}(0,1)$



L'aire hachurée illustre $\mathbb{P}(Z \leq z)$.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	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	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	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

Pour $z \geq 3$ on peut utiliser l'approximation

$$\mathbb{P}(Z \leq z) \approx 1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)\left(\frac{1}{z} - \frac{1}{z^3}\right).$$

Valeurs critiques au niveau 5% : loi normale centrée réduite

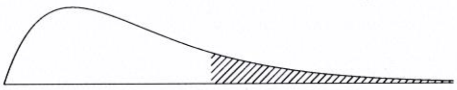
La valeur critique pour un test bilatéral de niveau 5% est $1,96$: $\mathbb{P}(|Z| \geq 1,96) = 0,05$.

La valeur critique pour un test unilatéral de niveau 5% est $1,645$: $\mathbb{P}(Z \geq 1,645) = 0,05$.

Valeurs critiques au niveau 5% loi du chi-carré

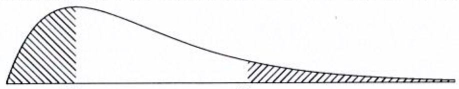
Soit Y une variable aléatoire de loi de chi-deux à ddl degrés de liberté. Les valeurs critiques pour effectuer un test unilatéral ou bilatéral au niveau 5% sont données au tableau suivant :

densité d'une loi du khi-deux à ddl degrés de liberté



L'aire hachurée illustre $\mathbb{P}(Y \geq y)$.

densité d'une loi du khi-deux à ddl degrés de liberté



L'aire hachurée illustre $\mathbb{P}(Y \leq y_1 \text{ ou } Y \geq y_2)$).

ddl	test unilatéral $\mathbb{P}(Y \geq y) = 0,05$	test bilatéral $\mathbb{P}(Y \leq y_1 \text{ ou } Y \geq y_2)) = 0,05$	
1	3,8414	0,0009	5,0238
2	5,9914	0,0506	7,3777
3	7,8147	0,2157	9,3484
4	9,4877	0,4844	11,1432
5	11,0704	0,8312	12,8325
6	12,5915	1,2373	14,4493
7	14,0671	1,6898	16,0127
8	15,5073	2,1797	17,5345
9	16,9189	2,7003	19,0227
10	18,3070	3,2469	20,4831
11	19,6751	3,8157	21,9200
12	21,0260	4,4037	23,3366

Fonction de répartition, sous $\mathcal{H}^0$, de la statistique de test à l'exercice 3

y	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0	0	0	0	0	0	0,0001	0,0001	0,0002	0,0004
1	0,0006	0,0008	0,0012	0,0016	0,0022	0,0029	0,0037	0,0046	0,0058	0,0070
2	0,0085	0,0102	0,0121	0,0142	0,0165	0,0191	0,0219	0,0250	0,0283	0,0319
3	0,0357	0,0398	0,0442	0,0488	0,0537	0,0589	0,0643	0,0700	0,0759	0,0821
4	0,0886	0,0953	0,1022	0,1094	0,1168	0,1245	0,1323	0,1404	0,1486	0,1571
5	0,1657	0,1745	0,1835	0,1926	0,2019	0,2113	0,2208	0,2305	0,2402	0,2501
6	0,2601	0,2701	0,2803	0,2904	0,3007	0,3110	0,3213	0,3317	0,3421	0,3525
7	0,3629	0,3733	0,3837	0,3941	0,4045	0,4148	0,4251	0,4354	0,4456	0,4557
8	0,4659	0,4759	0,4859	0,4958	0,5056	0,5154	0,5250	0,5346	0,5441	0,5534
9	0,5627	0,5719	0,5810	0,5899	0,5988	0,6075	0,6162	0,6247	0,6331	0,6414
10	0,6495	0,6575	0,6655	0,6733	0,6809	0,6885	0,6959	0,7032	0,7103	0,7174
11	0,7243	0,7311	0,7378	0,7443	0,7507	0,7570	0,7632	0,7692	0,7752	0,7810
12	0,7867	0,7923	0,7977	0,8031	0,8083	0,8134	0,8184	0,8233	0,8281	0,8328
13	0,8374	0,8419	0,8462	0,8505	0,8547	0,8587	0,8627	0,8666	0,8704	0,8741
14	0,8777	0,8812	0,8846	0,8880	0,8912	0,8944	0,8975	0,9005	0,9034	0,9063
15	0,9091	0,9118	0,9144	0,9170	0,9195	0,9219	0,9243	0,9266	0,9288	0,9310
16	0,9331	0,9352	0,9372	0,9391	0,9410	0,9429	0,9446	0,9464	0,9481	0,9497
17	0,9513	0,9528	0,9543	0,9558	0,9572	0,9586	0,9599	0,9612	0,9624	0,9636
18	0,9648	0,9660	0,9671	0,9682	0,9692	0,9702	0,9712	0,9721	0,9731	0,9739
19	0,9748	0,9756	0,9765	0,9772	0,9780	0,9787	0,9795	0,9801	0,9808	0,9815
20	0,9821	0,9827	0,9833	0,9839	0,9844	0,9849	0,9855	0,9859	0,9864	0,9869
21	0,9873	0,9878	0,9882	0,9886	0,9890	0,9894	0,9898	0,9901	0,9905	0,9908
22	0,9911	0,9914	0,9917	0,9920	0,9923	0,9926	0,9928	0,9931	0,9933	0,9936
23	0,9938	0,9940	0,9942	0,9944	0,9946	0,9948	0,9950	0,9952	0,9954	0,9955
24	0,9957	0,9959	0,9960	0,9961	0,9963	0,9964	0,9966	0,9967	0,9968	0,9969
25	0,9970	0,9971	0,9972	0,9973	0,9974	0,9975	0,9976	0,9977	0,9978	0,9979