

1 Exercices

1.1 Estimation d'une probabilité

Une pièce déséquilibrée tombe sur pile avec probabilité 0.2. On lance la pièce 100 fois et on compte le nombre de fois où la pièce tombe sur pile (i.e. le nombre de succès).

- Quelle est la loi de l'évènement "le i-eme lancé tombe sur pile" ? Rappeler son espérance et sa variance.
- Expliquer pourquoi le nombre de succès est une variable aléatoire ? Quelle est sa loi ?
- Calculer l'espérance et la variance de la fréquence de succès observée.
- Par quelle loi de probabilité continue peut-on approcher la loi Binomiale ? Utiliser cette approximation pour construire un intervalle de fluctuation au seuil 90% de la fréquence de succès.

1.2 Loi normale, fonction de répartition - vu en TD

Une entreprise fabrique des sacs poubelles. On note Y la variable aléatoire qui, à chaque sac prélevé au hasard dans la production, associe la masse maximale en kg qu'il peut supporter sans se déchirer. On suppose que $Y \sim \mathcal{N}(5, 0.4)$.

- On note F la fonction de répartition de la loi normale centrée réduite. Soit $Z \sim \mathcal{N}(5, 0.4)$. Pour tout $a < b$, exprimer $\mathbb{P}(a \leq Z \leq b)$ en fonction de $F(a)$ et $F(b)$.
- En déduire $\mathbb{P}(4.6 \leq Y \leq 5.4)$.
- Déterminer le réel h tel que $\mathbb{P}(Y \leq 5 + h) = 0.95$. Que représente le nombre h ?

1.3 Estimation sur une densité - vu en TD

Les individus d'une population possèdent un caractère X qui suit une loi de probabilité dont la densité est donnée par $f_\theta(x) = kx^2/\theta^3$ pour $x \in [0, \theta]$ et $f_\theta(x) = 0$ sinon. On s'intéresse au paramètre $\theta > 0$.

- Déterminer la constante k pour que f_θ soit une densité.
- On se donne (X_1, \dots, X_n) un échantillon de même loi que X . Montrer que la moyenne \bar{X}_n n'est pas un estimateur sans biais de θ .
- En déduire un estimateur sans biais de θ .
- Utiliser le tableau de valeurs suivant pour donner une estimation non biaisée de θ .

1.87	1.93	1.64	1.35	1.44	1.69	0.73	1.89	1.66	1.81
------	------	------	------	------	------	------	------	------	------

1.4 Estimation sur une densité inhabituelle

Soient X_1, \dots, X_n des variables aléatoires indépendantes ayant pour densité :

$$f_\theta(x) = \frac{1}{2}(1 + \theta x)\mathbb{I}_{[-1,1]}(x)$$

- Quelle contrainte doit-on imposer à θ pour que f_θ soit bien une densité de probabilité ?
- Par la méthode des moments, donner un estimateur $\hat{\theta}$ de θ .
- Peut-on dire que $\hat{\theta}$ est un bon estimateur de θ ?
- Pour $\theta = 0$ et $n = 100$, quelle est la probabilité que $\hat{\theta}$ soit inférieur à 0,5 ?

1.5 Estimation du paramètre d'une loi de Poisson

Le nombre de voitures passant chaque jour à un péage d'autoroute est modélisé par une variable de loi de Poisson. On a effectué un relevé pendant n jours et on veut utiliser ces données pour estimer le paramètre $\lambda > 0$ de la loi de Poisson. Rappelons que pour tout entier positif k , $\mathbb{P}(X_1 = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $\mathbb{E}[X_1] = \lambda$ et $\mathbb{E}[X_1^2] = \lambda + \lambda^2$.

1. La loi binomiale aurait-elle été pertinente pour modéliser ce phénomène ?
2. En déduire deux estimateurs de λ avec la méthode des moments.
3. Déterminer leur biais.

1.6 Compromis biais/variance (assez ambitieux)

Pour estimer la moyenne μ d'une population on utilise souvent la moyenne \bar{X}_n de l'échantillon. Pour réduire la variance de \bar{X}_n , on envisage d'utiliser l'estimateur $a\bar{X}_n$, avec $0 \leq a \leq 1$.

1. Calculer le biais au carré et la variance de $a\bar{X}_n$.
2. Faire un graphique du biais au carré et de la variance de $a\bar{X}_n$ en fonction de a .
3. Qu'observe-t-on ?
4. Pour quelle valeur de a obtiendrait-on un bon compromis entre le biais au carré et la variance ?

1.7 Intervalle de confiance et interprétations

Un fabricant de voitures fait un test d'efficacité d'un certain modèle. Il mesure les litres de consommation d'essence pour 100 kilomètres sur 13 de ses véhicules et obtient une moyenne de 14,29. On suppose que l'on connaît la variance σ^2 de la population, $\sigma^2 = 16$.

1. Donner un intervalle de confiance au niveau 95% pour la moyenne.
2. Quel niveau le fabricant devrait-il choisir pour obtenir un intervalle de longueur 3 ?
3. Combien de véhicules supplémentaires devrait-il tester pour obtenir un intervalle de confiance au niveau 99% de longueur 2 ?

1.8 Intervalle de confiance dans un cadre gaussien exotique

On voudrait estimer la vitesse moyenne μ à laquelle est parcouru le contournement de Nice. Les variables aléatoires X_1, \dots, X_n représentent les mesures effectuées sur un échantillon pris uniquement parmi les hommes. On suppose qu'elles proviennent d'une loi normale d'espérance inconnue μ et de variance connue σ_H^2 . En deuxième lieu, on récolte un échantillon de mesures de même taille uniquement parmi les femmes Y_1, \dots, Y_n . Ces mesures proviennent d'une loi normale d'espérance inconnue μ et de variance connue σ_F^2 . Pour estimer μ , on considère l'estimateur :

$$\hat{\mu} = \frac{1}{2} (\bar{X} + \bar{Y}).$$

1. Donner la loi de $\hat{\mu}$.
2. Donner un intervalle de confiance à 95% pour μ basé sur $\hat{\mu}$.

1.9 Construction d'un intervalle de confiance pour la loi exponentielle

Chaque jour, un avion subit un retard aléatoire au départ, évalué en minutes, indépendant des retards des autres jours et dont la loi est supposée exponentielle de paramètre λ . Sur une année (365 jours), le retard moyen a été de 8 minutes.

1. Rappeler l'espérance et la variance de cette loi.
2. Construire un intervalle de confiance de niveau 95% pour λ .

1.10 Test d'hypothèse sur la moyenne d'une population

On considère le test d'hypothèse suivant:

$$\mathcal{H}_0 : \mu = 18 \quad (1)$$

Sachant que l'écart type $\sigma = 4.5$, un échantillon de taille $n = 48$ fournit une moyenne d'échantillon de $\bar{x}_n = 17$.

1. Quelle est la loi de la statistique de test?
2. Calculez la valeur de la statistique de test?
3. Quelle est la valeur de la "valeur p" (p-valeur)?
4. Donner la règle de rejet pour le test, que peut-on conclure?
5. Au seuil de signification 5% et 1% quelle est votre conclusion?

1.11 Test d'hypothèse sur une proportion

Tester si la probabilité de s'inclure après remboursement dans le système bancaire classique dépasse 33%. Pour ce faire, on observe sur un échantillon de 50 clients que 22 clients deviennent inclus.

1. Soit X la variable aléatoire désignant l'état d'un client (inclus dans le système bancaire ou pas). Quelle est la loi de X ?
2. Effectuer un test d'hypothèse.
3. l'hypothèse nulle est-elle rejetée au seuil 5%?

1.12 Test d'hypothèse: moyenne d'une population, variance inconnue

On considère le test d'hypothèse suivant, où μ désigne la moyenne d'une population:

$$\mathcal{H}_0 : \mu \geq 34 \quad (2)$$

On sait que la population a une distribution normale. Un échantillon de taille $n = 50$ fournit une moyenne d'échantillon de 31 et un écart-type d'échantillon de 5.

1. Déterminer la statistique de test, et calculer sa valeur.
2. Au seuil de signification $\alpha = 0.05$ quelle est votre conclusion?

1.13 Test d'hypothèse sur une proportion 2

On considère le test suivant où p désigne la proportion d'une population:

$$\mathcal{H}_0 : p = 0.6 \quad (3)$$

1. Un échantillon de taille $n = 200$ a fournit une proportion d'échantillon de 0.67. Calculer la valeur p et conclure, pour un seuil de signification $\alpha = 0.05$
2. Même question pour une proportion d'échantillon de 0.62. Puis de 0.57

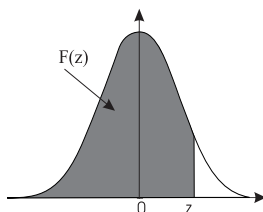
1.14 Application: Test d'hypothèse 1

On s'intéresse au problème des algues toxiques qui atteignent certaines plages de France. Après étude, on constate que 10% des plages sont atteintes par ce type d'algues et on veut tester l'influence de rejets chimiques sur l'apparition de ces algues. Pour cela 50 plages proches de zones de rejets chimiques sont observées. On constate que 10 plages sont atteintes par l'algue. Pouvez-vous répondre à la question: "les rejets chimiques ont-ils modifié de façon significative avec le risque $\alpha = 5\%$ le nombre de plages atteintes?"

1.15 Application: Test d'hypothèse 2

Un appareil de télécommunication reçoit un signal stocké à chaque petite unité de temps dans une suite de variables (X_n) . Cet appareil doit détecter un signal effectif en le différenciant d'un bruit. On suppose que le bruit est une suite de variables indépendantes de loi normale de moyenne nulle et de variance 1. Pour un signal, la moyenne n'est pas nulle. Aujourd'hui, on a observé une suite de 40 variables (x_1, \dots, x_{40}) , supposées indépendantes de variance 1. La moyenne empirique vaut 0.6. S'agit-il de bruit? Répondre en construisant un test.

Fonction de répartition de la loi normale centrée réduite
(probabilité $F(z)$ de trouver une valeur inférieure à z)



z	0,0	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

Table pour les grandes valeurs de z

z	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
F(z)	0,998650	0,999032	0,999313	0,999517	0,999663	0,999767	0,999841	0,999892	0,999928	0,999952
z	4,0	4,1	4,2	4,3	4,4	4,5	4,6	4,7	4,8	4,9
F(z)	0,999968	0,999979	0,999987	0,999991	0,999995	0,999997	0,999998	0,999999	0,999999	1,000000

Nota. La table donne $F(z)$ pour z positif. Pour z négatif, il faut prendre le complément à l'unité de la valeur lue dans la table. Exemple : $F(-1,37) = 1 - F(1,37) = 1 - 0,9147 = 0,0853$.

2 Remarques/Éléments de correction

Section 1.4

1. compris entre -1 et 1 car on veut un truc positif
2. $3 \bar{X}_n$
3. biais=0 et variance= $\frac{3-\theta^2}{n}$, oui bon estimateur
4. utilisation du TCL + lecture dans la table de la $N(0,1)$

Section 1.5

1. non, aucune raison de poser un N maximal a priori
2. utilisation de l'estimateur \bar{s}_n^2 de la variance
3. on pourra donner la valeur de l'espérance de \bar{s}_n^2

Section 1.6

1. $B^2 = \mu^2(a-1)^2$ et $Var = \frac{\sigma^2}{n} a^2$.
2. ...
3. on augmente le biais (en valeur absolue) quand a décroît
4. on peut chercher le a qui minimise la somme $B^2 + Var$, on trouve $\frac{\mu^2}{\mu^2 + \frac{\sigma^2}{n}}$

Section 1.7

1. formule habituelle $\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}q, \bar{X}_n + \frac{\sigma}{\sqrt{n}}q \right]$, ce serait bien de leur faire retrouver
2. longueur= $2q \frac{\sigma}{\sqrt{n}}$, on veut donc $q \approx 1,35$, alors $\alpha \approx 82,2\%$.
3. $n = (q_{0,995} \times \sigma)^2 = (2,576 \times 4)^2 \approx 106$

Section 1.8

1. en utilisant l'indépendance, on montre que $\hat{\mu}$ est de loi $N(\mu, \frac{\sigma_H^2 + \sigma_F^2}{4n})$
2. $\left[\hat{\mu} - 1,96 \sqrt{\frac{\sigma_H^2 + \sigma_F^2}{4n}}, \hat{\mu} + 1,96 \sqrt{\frac{\sigma_H^2 + \sigma_F^2}{4n}} \right]$

Section 1.9

1. peut-être leur faire écrire la formule de l'espérance et leur demander l'idée du calcul
2. TCL, on arrive à $\left[\frac{1}{\bar{X}_n} \left(1 - \frac{1,96}{\sqrt{n}} \right), \frac{1}{\bar{X}_n} \left(1 + \frac{1,96}{\sqrt{n}} \right) \right] = [0,112, 0,138]$

3 Questions de cours

3.1 Fonction de répartition, quantiles

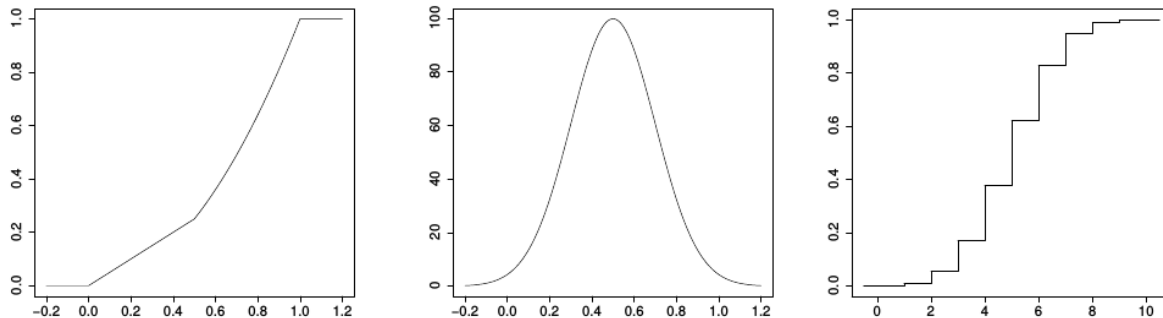
1. Qu'est ce que le premier décile de la loi d'une variable aléatoire ? Qu'est ce que le q -ème quantile ?
2. Déterminer la médiane et le quantile d'ordre 75% de la loi exponentielle de paramètre λ . **on pourra donner la formule de la fct de répartition dans ce cas**

3.2 Centrer, réduire et fonctions de répartition

Soit X une variable aléatoire de loi normale, de moyenne m et de variance σ^2 .

1. On pose $Y = \frac{X-m}{\sigma}$. Donner la loi de Y .
2. Faire le lien entre les fonctions de répartition de X et de Y .

3.3 Fonction de répartition, loi discrète ou continue



1. Dire des graphiques précédents si ils correspondent à une fonction de répartition.
2. Lesquels correspondent à une loi discrète ? À une loi continue ?

3.4 Fonction de répartition, loi discrète ou continue (bis)

1. Citer deux lois discrètes et deux lois continues.
2. À partir de la fonction de répartition, peut-on déduire si une loi est discrète ou continue ?

3.5 Espérances et variances

Soit $(X_i)_{1 \leq i \leq n}$ une suite de variables aléatoires i.i.d.

1. Rappeler les valeurs de l'espérance et de la variance de l'estimateur \bar{X}_n .
2. Préciser ces valeurs dans le cas où les X_i sont de loi de Bernoulli.

3.6 Bon estimateur pour la loi de Bernoulli

On dispose d'un n-échantillon $(X_i)_{1 \leq i \leq n}$ de loi de Bernoulli de paramètre p .

1. À quelle(s) condition(s) un estimateur \hat{p} est-il un bon estimateur de p ?
2. Parmi les estimateurs $\hat{p}_1 = \frac{X_1 + X_2}{2}$, $\hat{p}_2 = \bar{X}_n$ et $\hat{p}_3 = \bar{X}_n - \frac{1}{n}$, lesquels sont bons ?