

TD 4 – Echantillonnage

Exercice 1. (*théorème Central Limit en l'absence de modèle*)

Les français gagnent 2000 € par mois en moyenne et la variance de leurs salaires vaut 1000 €².

Le salaire x d'un français choisi au hasard est une variable aléatoire d'espérance $\mu = 2000$, de variance $\sigma^2 = 1000$, dont on ignore la loi de probabilité $\mathcal{L}(x)$.

Les salaires $x_1, \dots, x_n$ de n français choisis au hasard constituent un échantillon aléatoire de $\mathcal{L}(x)$. Ainsi, la moyenne des salaires de l'échantillon : $\bar{x} = \sum_{i=1}^n x_i/n$ est approximativement distribuée, quand n est grand, selon $\mathcal{N}(2000; 1000/n)$, d'après TCL.

1. Avec quelle probabilité le salaire moyen parmi trente français est-il supérieur à 2010 € ?

$$n = 30 \text{ et } \bar{x} \underset{\text{ap.}}{\sim} \mathcal{N}(2000; 1000/30).$$

$$\mathbb{P}(\bar{x} > 2010) = \mathbb{P}(\mathcal{N}(2000; 1000/30) > 2010) = 1 - \phi((2010 - 2000)/\sqrt{1000/30}) = 1 - \phi(\sqrt{3}) \approx 1 - 0,96 = 0,04.$$

Il y a 4% de chances que le salaire moyen parmi trente français soit supérieur à 2010 euros.

2. Quelle est la taille d'un échantillon de français dont le salaire moyen est supérieur à 2010 € avec 2% de chances ?

$$\text{On cherche } n \text{ de sorte que : } \mathbb{P}(\bar{x} > 2010) = 0,02 ; \mathbb{P}(\mathcal{N}(2000; 1000/n) > 2010) = 0,02 ; 1 - \phi((2010 - 2000)/\sqrt{1000/n}) = 0,02 ; \phi((2010 - 2000)/\sqrt{1000/n}) = 0,98 ; (2010 - 2000)/\sqrt{1000/n} = \phi^{-1}(0,98) ; \sqrt{n/10} = \phi^{-1}(0,98) ; n = 10 \times (\phi^{-1}(0,98))^2 \approx 43.$$

Il faut interroger 43 français pour que la moyenne des salaires ait 2% de chances de dépasser 2010 euros.

Exercice 2. (*théorème Central Limit et loi de Bernoulli*)

70% des étudiants vivent en ville.

n étudiants sont choisis au hasard ; la variable x_i indique si l'étudiant i vit en ville ($x_i = 1$) ou à la campagne ($x_i = 0$). Ainsi, $x_1, \dots, x_n$ est un échantillon aléatoire de la loi de Bernoulli de paramètre $p = 0,7$; les variables $x_1, \dots, x_n$ ont la même espérance : $\mu = 0,7$ et la même variance : $\sigma^2 = 0,7 \times 0,3 = 0,21$.

$\bar{x} = \sum_{i=1}^n x_i/n$ est la proportion des étudiants citadins dans l'échantillon ; quand n est grand, $\bar{x}$ est approximativement distribuée selon $\mathcal{N}(0,7; 0,21/n)$, d'après TCL.

1. Avec quelle probabilité un échantillon de cent étudiants contient-il au moins 65% de citadins ?

$$n = 100 \text{ et } \bar{x} \underset{\text{ap.}}{\sim} \mathcal{N}(0,7; 0,21/100).$$

$$\mathbb{P}(\bar{x} \geq 0,65) = \mathbb{P}(\mathcal{N}(0,7; 0,21/100) \geq 0,65) = 1 - \phi((0,65 - 0,7)/\sqrt{0,21/100}) = 1 - \phi(-0,5/\sqrt{0,21}) = \phi(0,5/\sqrt{0,21}) \approx 0,86.$$

Il y a 86% de chances qu'un échantillon de cent étudiants comporte au moins 65% de citadins.

2. Quelle est la taille d'un échantillon contenant au moins 65% de citadins avec neuf chances sur dix ?

$$\text{On cherche } n \text{ de sorte que : } \mathbb{P}(\bar{x} \geq 0,65) = 0,9 ; \mathbb{P}(\mathcal{N}(0,7; 0,21/n) \geq 0,65) = 0,9 ; 1 - \phi((0,65 - 0,7)/\sqrt{0,21/n}) = 0,9 ; \phi(0,05\sqrt{n/0,21}) = 0,9 ; 0,05\sqrt{n/0,21} = \phi^{-1}(0,9) ; n = 0,21 \times (\phi^{-1}(0,9)/0,05)^2 \approx 138.$$

Il faut interroger 138 étudiants pour que la proportion de citadins ait 90% de chances de dépasser 0,65.

Exercice 3. (*théorème Central Limit et loi binomiale*)

Le nombre d'absences mensuelles d'un salarié est distribué selon une loi binomiale de paramètres $n = 20$ et $p = 0,1$.

n salariés sont choisis au hasard ; x_i indique le nombre d'absences mensuelles du salarié i . Ainsi, $x_1, \dots, x_n$ est un échantillon aléatoire de la loi binomiale $b(20; 0,1)$; $x_1, \dots, x_n$ ont la même espérance : $\mu = 20 \times 0,1 = 2$ et la même variance : $\sigma^2 = 20 \times 0,1 \times 0,9 = 1,8$.

$\bar{x} = \sum_{i=1}^n x_i/n$ est le nombre moyen d'absences calculé sur n salariés ; quand n est grand, $\bar{x}$ est approximativement distribuée selon $\mathcal{N}(2; 1,8/n)$, d'après TCL

1. Avec quelle probabilité le nombre moyen d'absences observé parmi cinquante salariés vaut-il au plus 2,1 ?

$$n = 50 \text{ et } \bar{x} \underset{\text{ap.}}{\sim} \mathcal{N}(2; 1,8/50).$$

$$\mathbb{P}(\bar{x} \leq 2,1) = \mathbb{P}(\mathcal{N}(2; 1,8/50) \leq 2,1) = \phi((2,1 - 2)/\sqrt{1,8/50}) \approx 0,7.$$

Parmi 50 salariés, le nombre moyen d'absences en un mois vaut au plus 2,1 avec 70% de chances.

2. Quelle est la taille d'un échantillon de salariés dans lequel le nombre moyen d'absences mensuelles est supérieur à 1,9 avec 95% de chances ?

On cherche n de sorte que : $\mathbb{P}(\bar{x} > 1,9) = 0,95$; $\mathbb{P}(\mathcal{N}(2; 1,8/n) > 1,9) = 0,95$; $1 - \phi((1,9 - 2)/\sqrt{1,8/n}) = 0,95$; $\phi(0,1\sqrt{n/1,8}) = 0,95$; $0,1\sqrt{n/1,8} = \phi^{-1}(0,95)$; $n = 1,8 \times (\phi^{-1}(0,95)/0,1)^2 \approx 487$.

Il faut interroger 487 salariés pour que le nombre moyen d'absences mensuelles soit supérieur à 1,9 avec 95% de chances.

Exercice 4. (théorème Central Limit et loi de Poisson)

Le nombre de pannes hebdomadaires du tram bordelais est distribué selon une loi de Poisson de paramètre $\lambda = 10$. Avec quelle probabilité y aura-t-il au plus onze pannes par semaine en moyenne en 2022 ?

n semaines d'observation sont choisies ; la variable x_i indique le nombre de pannes du tram dans la semaine i . Ainsi, $x_1, \dots, x_n$ est un échantillon aléatoire de la loi de Poisson $\mathcal{P}(10)$; les variables $x_1, \dots, x_n$ ont la même espérance : $\mu = 10$ et la même variance : $\sigma^2 = 10$.

$\bar{x} = \sum_{i=1}^n x_i/n$ est le nombre moyen de pannes hebdomadaires, calculé sur n semaines ; quand n est grand, $\bar{x}$ est approximativement distribuée selon $\mathcal{N}(10; 10/n)$, d'après TCL.

Comme 2022 compte cinquante-deux semaines ($n = 52$), le nombre moyen $\bar{x}$ de pannes par semaine, cette année là est distribué selon $\mathcal{N}(10; 10/52)$. $\mathbb{P}(\bar{x} \leq 11) = \mathbb{P}(\mathcal{N}(10; 10/52) \leq 11) = \phi((11 - 10)/\sqrt{10/52}) = \phi(\sqrt{5}, 2) \approx 0,99$.

Il y a 99% de chances que le tram bordelais connaisse au plus onze pannes par semaines en moyenne au cours de l'année 2022.

Exercice 5. (théorème Central Limit et loi exponentielle)

1. Avec quelle probabilité la moyenne d'un 100-échantillon issu d'une loi exponentielle de paramètre $\lambda = 0,01$ est-elle comprise entre 95 et 105 ?

$x_1, \dots, x_n$ est un n -échantillon de la loi $\mathcal{E}(0,01)$; les variables $x_1, \dots, x_n$ ont la même espérance : $\mu = 100$ et la même variance : $\sigma^2 = 10000$. Quand n est grand, $\bar{x} = \sum_{i=1}^n x_i/n$ est approximativement distribuée selon $\mathcal{N}(100; 10000/n)$, d'après TCL.

Si $n = 100$, alors $\bar{x} \underset{\text{ap.}}{\sim} \mathcal{N}(100; 100)$.

$\mathbb{P}(95 \leq \bar{x} \leq 105) = \mathbb{P}(95 \leq \mathcal{N}(100; 100) \leq 105) = \phi((105 - 100)/\sqrt{100}) - \phi((95 - 100)/\sqrt{100}) = \phi(0,5) - \phi(-0,5) = 2\phi(0,5) - 1 \approx 1 \times 0,69 - 1 = 0,38$.

Il y a 38% de chances que la moyenne d'un 100-échantillon d'une loi exponentielle de paramètre 0,01 soit comprise entre 95 et 105.

2. Ecrivez l'énoncé d'un problème auquel 1. apporterait une solution.

$\mathcal{E}(0,01)$ peut modéliser la durée de vie d'une tortue dans une espèce où l'espérance de vie est proche de 100. Ainsi, 1. permettrait d'affirmer : il y a 38% de chances que la durée de vie moyenne de cent tortues soit comprise entre 95 et 105 ans.

Exercice 6. (loi normale (et théorème Central Limit ?))

La durée de la course (exprimée en heures) d'un participant au dernier marathon de New-York est distribuée selon une loi normale de moyenne $\mu = 4$ et de variance $\sigma^2 = 1,2$. Avec quelle probabilité la durée de course moyenne de trente coureurs dépasse-t-elle 4 h 06 mins ?

n marathonien sont choisis au hasard. x_i désigne la durée de course (en heures) du coureur i . Ainsi, $x_1, \dots, x_n$ est un n -échantillon de la loi normale d'espérance : $\mu = 4$ et de variance $\sigma^2 = 1,2$. Quelle que soit la valeur de n , $\bar{x} = \sum_{i=1}^n x_i/n$ est distribuée selon $\mathcal{N}(4; 1,2/n)$. Ainsi, $\mathcal{N}(4; 1,2/n)$ est la loi exacte de $\bar{x}$ quelle que soit la taille de l'échantillon. Ce résultats, plus fort que la conclusion à laquelle conduit le théorème Central Limit, est une conséquence de l'additivité de la loi normale. En effet, puisque $x_1, \dots, x_n$ sont indépendantes, $\sum_{i=1}^n x_i$ est distribuée selon $\mathcal{N}(n \times 4; n \times 1,2)$ et $\sum_{i=1}^n x_i/n$ selon la loi normale de moyenne $n \times 4/n = 4$ et de variance $n \times 1,2/n^2 = 1,2/n$.

Si $n = 30$, alors $\bar{x} \sim \mathcal{N}(4; 1,2/30)$ et $\mathbb{P}(\bar{x} \geq 4,1) = \mathbb{P}(\mathcal{N}(4; 1,2/30) \geq 4,1) = 1 - \phi((4,1 - 4)/\sqrt{1,2/30}) = 1 - \phi(0,5) \approx 0,31$. Il y a 31% de chances que la durée de course moyenne calculée sur trente marathonien soit supérieure ou égale à quatre heures et six minutes.

Exercice 7. (théorème Central Limit en l'absence de modèle)

En Italie, le salaire annuel moyen est de 36 k€ et la variance des salaires vaut 100 k€².

1. Avec quelle probabilité le salaire moyen est-il inférieur à 37 k€ dans un échantillon de trente italiens ?

Même méthode que pour Exercice 1. Il y a 71% de chances que le salaire moyen parmi trente italiens soit inférieur ou égal à 37 k€.

2. Combien d'italiens doit-on sonder pour que le salaire moyen dans l'échantillon soit supérieur à 35 k€ avec 98% de chances ?

Il faut interroger 422 italiens pour que la moyenne de leurs salaires ait 98% de chances de dépasser 35 k€.

Exercice 8. (*théorème Central Limit et loi de Bernoulli*)

60% des français aiment le chocolat.

1. Avec quelle probabilité y-a-t-il au moins 58% d'amateurs de chocolat parmi les 400 invités au meeting de M. Li ?

Même méthode que pour Exercice 2. Il y a 79% de chances que la proportion d'amateurs de chocolats parmi quatre cents invités soit supérieure ou égale à 0,58.

2. Combien de personnes M. Li doit-il inviter à son meeting pour avoir au moins 55% d'amateurs de chocolat avec 95% de chances ?

Il faut inviter 260 personnes pour que la proportion d'amateurs de chocolat dépasse 0,55 avec 95% de chances.
