# TD 2 - Intervalles de Confiance

**Exercice 2.1**. On effectue un sondage sur un échantillon de 400 électeurs. On relève 212 intentions de vote en faveur d'un candidat $A$, 188 en faveur de $B$.

1. Donner au niveau de 95%, un intervalle de confiance asymptotique des intentions de vote en faveur de $A$ dans la population entière.

2. Quelle taille minimale de l'échantillon faudrait-il prendre pour que, au même niveau de 95% et pour la même proportion de votants en faveur de $A$, l'intervalle ne contienne pas la valeur 0.5 ?

**Solution:**

1. Soit $\hat{p}_n$, la variable aléatoire correspondant à la proportion d'individus votant pour $A$ dans un échantillon de taille $n$, et soit $p$ la proportion théorique dans la population.

$\hat{p}_n$ est la moyenne empirique de $n$ variables i.i.d de Bernoulli $B(p)$ (nous rappelons que $B(p)$ a pour moyenne $p$ et pour variance $p(1-p)$. Donc, le théorème central limite donne:

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{\mathcal{L}} \mathcal{N}(0, p(1-p))$$

La loi limite dépend du paramètre inconnu, mais on peut réécrire:

$$\sqrt{n}\frac{(\hat{p}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

$\sqrt{n}\frac{(\hat{p}_n - p)}{\sqrt{p(1-p)}}$ est donc une variable aléatoire asymptotiquement pivotale. Malgré tout, la construction de l'IC nous conduit à résoudre des inégalités. C'est encore possible explicitement ici, mais il y a une méthode plus simple, en utilisant le théorème de Slutsky. On a:

$$\sqrt{n}\frac{(\hat{p}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

et comme $\hat{p}_n \xrightarrow{\mathbb{P}} p$ par la loi des grands nombres, en utilisant le théorème de continuité

$$\frac{\sqrt{p(1-p)}}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} \xrightarrow{\mathbb{P}} 1.$$

On peut donc utiliser le théorème de Slutsky pour conclure que:

$$\sqrt{n}\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n(1-\hat{p}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Cette variable aléatoire est donc également asymptotiquement pivotale, mais plus simple que la précédente. On en déduit un intervalle de confiance de niveau asymptotique $1-\alpha$:

$$I = [\hat{p}_n - q^{\mathcal{N}}_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + q^{\mathcal{N}}_{1-\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}]$$

Numériquement, on trouve $\hat{p}_{400} = 212/400 = 0.53, q^N_{0.975} = 1.96$ et $\sqrt{\frac{\hat{p}_{400}(1-\hat{p}_{400})}{400}} \simeq 2.510^{-2}$. D'où $I = [0.48,\ 0.58\ ]$.

N.B.: On a la loi exacte pour le nombre d'invidus $N_A$ qui votent pour $A$ dans l'échantillon, $N_A$ suit une loi binomiale de paramètres $N, p$: $N_A \sim B(N, p)$. Malheureusement, il n'est pas possible de déterminer à partir de cette loi exacte une fonction pivotale. C'est pour cette raison que nous utilisons une fonction pivotale asymptotique.

2. On résout $0.53 - 1.96\sqrt{\frac{053(1-053)}{n}} > 0.5$ et on obtient $n > (1.96/0.03)^2 \times 0.53 \times 0.47 \simeq 1063$. D'où $n = 1064$ convient.

**Exercice 2.2**. 1. Soit $X_1, \cdots, X_N$ un échantillon i.i.d. pour la loi de Poisson $\mathcal{P}(\theta)$. Donner l'estimateur de $\theta$ par la méthode des moments et montrer qu'il est asymptotiquement normal. Trouver une transformation qui stabilise la variance asymptotique associée à l'estimateur, c'est à dire une fonction $h$ de classe $C^1$ sur l'ensemble des paramètres telle que la variance asymptotique associée à l'estimateur de $h(\theta)$ ne dépende pas de $\theta$.

2. On dispose d'un échantillon i.i.d. de taille $N = 400$ d'une loi de Poisson $\mathcal{P}(\theta)$ de paramètre $\theta$ inconnu. Proposer deux intervalles de confiance au niveau asymptotique 0.99 pour $\theta$, fondés sur l'estimateur du maximum de vraisemblance. Le premier fera appel au résultat de la question 1., le second au théorème de Slutsky.

**Solution:** 1. $\mathbb{E}(X_1) = \theta$. L'estimateur de la méthode des moments est donc $\hat{\theta} = \overline{X}$. De plus, on a: $\mathbb{V}(X_1) = \theta$. Le théorème central limite donne: $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta)$. L'estimateur est asymptotiquement normal de variance associée $\theta$. Comme $\theta$ est inconnu, on cherche à se ramener à un estimateur dont la variance asymptotique associée ne dépend pas de $\theta$.

Soit $h$ de classe $C^1$ telle que $h'(\theta) \neq 0$. Par la méthode delta, nous avons:

$$\sqrt{N}(h(\hat{\theta}) - h(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta(h'(\theta))^2)$$

En prenant $h(\theta) = \sqrt{\theta}$, nous avons alors:

$$\sqrt{N}(\sqrt{\hat{\theta}} - \sqrt{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{1}{4})$$

2. On trouve l'intervalle de confiance asympotique de niveau $0.99 = 1 - \alpha$ (donc $\alpha = 0.01$ :

$$I_\infty = [(\max(0,\ \sqrt{\overline{x}} - \frac{q^{\mathcal{N}}_{1-\alpha/2}}{2\sqrt{N}}))^2,\ (\sqrt{\overline{x}} + \frac{q^{\mathcal{N}}_{1-\alpha/2}}{2\sqrt{N}})^2].$$

En utilisant la table de la loi normale, on trouve $q^{\mathcal{N}}_{1-\alpha/2} = 2.58$. L'application numérique donne :

$$I_\infty = [(\max(0,\ \sqrt{\overline{x}} - 0,065))^2,\ (\sqrt{\overline{x}} + 0,065)^2]\ .$$

Une autre solution possible consiste à utiliser le fait que : $\dfrac{\sqrt{N}(\overline{x}-\theta)}{\sqrt{\overline{x}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ d'après le théorème de Slutsky. Il fournit un intervalle de confiance légèrement différent mais asympotiquement équivalent au précedent :

$$I'_\infty = [\max\left(0, \overline{x} - \frac{q^{\mathcal{N}}_{1-\alpha/2}}{\sqrt{N}}\sqrt{\overline{x}}\right),\ \overline{x} + \frac{q^{\mathcal{N}}_{1-\alpha/2}}{\sqrt{N}}\sqrt{\overline{x}}]$$

On reconnaît le développement des carrés dans $I_\infty$ sans le terme en $1/N$ asymptotiquement négligeable devant les deux autres termes. Numériquement

$$I'_\infty = [\max\left(0, \overline{x} - 0.129\sqrt{\overline{x}}\right),\ \overline{x} + 0.129\sqrt{\overline{x}}]\ .$$

**Exercice 2.3**. On considère un échantillon $X_1, \cdots, X_n$ de variables aléatoires i.i.d. selon une loi géométrique décrite par sa fonction de masse

$$\forall k \in \mathbb{N}^*,\ p_\theta(k) = \mathbb{P}_\theta(X = k) = \theta(1-\theta)^{k-1}$$

où $\theta$ est le paramètre inconnu de la loi. On admettra les relations

$$\mathbb{E}_\theta(X) = \frac{1}{\theta} \text{ et } \mathbb{V}_\theta(X) = \frac{1-\theta}{\theta^2}$$

1. Trouver l'estimateur du maximum de vraisemblance de $\theta$, que l'on notera $\hat{\theta}_n$.
2. Appliquer, en le justifiant, le théorème central limite à la moyenne empirique $\overline{X}_n$.
3. En déduire, en utilisant le théorème de Slutsky, une fonction asymptotiquement pivotale avec une dépendance en $\theta$ aussi simple que possible.
4. Construire un intervalle de confiance asymptotique pour $\theta$ de niveau $1-\alpha$.
5. Application numérique: On observe un échantillon de taille $n = 25$ et on obtient la distribution empirique suivante :

| *Valeurs de k* | *1* | *2* | *3* | *4* | *5* | *6* | *7* |
|---|---|---|---|---|---|---|---|
| *Nombre d'occurrences $n_k$* | *11* | *4* | *3* | *3* | *2* | *1* | *1* |

Calculer l'intervalle de confiance sur $\theta$ pour ce jeu de données au niveau 0,95.

**Solution:**

1. Soit l'échantillon d'observations $x = (x_1, \cdots, x_n)$. Nous avons la vraisemblance:

$$\mathcal{L}(\theta; x) = \prod_{i=1}^{n} p_\theta(x_i) = \theta^n (1-\theta)^{n\overline{x}-n}$$

La $\log$-vraisemblance :

$$\ell(\theta) = n\ln\theta + n(\overline{x}-1)\ln(1-\theta)\ .$$

Si $\overline{x}=1$, pas de maximum, sinon:

$$\ell'(\theta) = n/\theta - n(\overline{x}-1)/(1-\theta)$$

et l'EMV est défini par la CN, $\ell'(\theta)=0$ soit $\hat{\theta}=1/\overline{x}$
Justification : $\forall\theta, \ell''(\theta)<0$ $l$ est concave, donc $\hat{\theta}$ est bien un max.
2. On a, d'après le TCL:

$$\sqrt{n}(\overline{X}_n - 1/\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0,(1-\theta)/\theta^2)$$

car $\overline{X}_n$ moyenne empirique de v.a. i.i.d. telles que $\mathbb{E}(X_1^2)<+\infty$.
3. Par la loi des grands nombres, on a $\overline{X}_n \xrightarrow{\mathbb{P}} \frac{1}{\theta}$.
donc par le théorème de continuité $\sqrt{1-1/\overline{X}_n} \xrightarrow{\mathbb{P}} \sqrt{1-\theta}$
et enfin avec le théorème de Slutsky

$$\sqrt{n}\frac{(\theta\overline{X}_n-1)}{\sqrt{1-(1/\overline{X}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

4. On pose $\Phi$ la fonction de répartition de la loi $\mathcal{N}(0,1)$. En utilisant le TCL, on a

$$\lim_{n\to\infty}\left[\mathbb{P}(\Phi^{-1}(\alpha/2) \le \sqrt{n}\frac{(p\overline{X}-1)}{\sqrt{1-(1/\overline{X}_n)}} \le \Phi^{-1}(1-\alpha/2))\right] = 1-\alpha$$

et on a:

$$I_\infty(n,\,\alpha) = \left[\frac{1}{\overline{x}}\left((1+\frac{\sqrt{1-(1/\overline{x})}}{\sqrt{n}}\Phi^{-1}(\alpha/2)\right), \frac{1}{\overline{x}}\left((1+\frac{\sqrt{1-(1/\overline{x})}}{\sqrt{n}}\Phi^{-1}(1-(\alpha/2))\right)\right]$$

A noter que: $\Phi^{-1}(1-(\alpha/2)) = q^{\mathcal{N}}_{1-\alpha/2}$ et $\Phi^{-1}(\alpha/2) = -q^{\mathcal{N}}_{1-\alpha/2}$.

5. A.N. $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{7} kn_k = 2.52.$

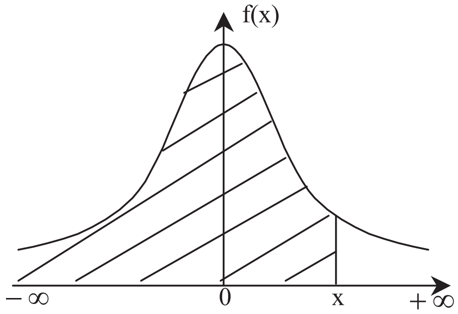
$$\Phi^{-1}(0,975) = 1.96,\ I_\infty(25, 0.05) = [0.27; 0.51].$$

Malgré tout, avec $N=25$, l'approximation asymptotique reste discutable.

# Annexe: Table

## Loi Normale centrée réduite

Probabilité de trouver une valeur inférieure à x.




$$F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du$$

| X | 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 |
| **3,0** | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |
| **3,1** | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 | 0,9993 |
| **3,2** | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 | 0,9995 |
| **3,3** | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9997 |
| **3,4** | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9998 |
| **3,5** | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 |