Simulation d'un sondage. Fourchettes d'échantillonnage et intervalles de confiance.

Michel Henry(*)

I. Situations de sondages

La pratique des sondages s'est largement développée au cours des trente dernières années, mettant à profit la diffusion des outils informatiques. Les méthodes de sondage se sont diversifiées et affinées. Parmi elles, seules les méthodes aléatoires permettent de garantir une approximation déterminée sur les résultats affichés, à un niveau de confiance souhaité. Quelle que soit leur complexité, elles reposent toutes sur le même principe que les sondages aléatoires simples. L'estimation d'un pourcentage à partir d'un échantillon aléatoire découle directement de la loi des grands nombres, obtenue dans sa forme la plus élémentaire par Jacques Bernoulli au début du 18e siècle.

Formalisons un peu une telle situation de sondage. Si, dans une population statistique, des éléments en proportion p inconnue présentent une certaine propriété M, on peut avoir une indication sur la valeur de p en dénombrant les éléments qui ont la propriété M dans un échantillon aléatoire de taille n assez grande. Pour simplifier, on suppose que la population est assez vaste pour pouvoir considérer le prélèvement de cet échantillon comme non exhaustif (i.e. « avec remises »). Par exemple, dans les sondages d'opinions, M peut être le choix préférentiel d'un consommateur ou d'un électeur.

Le fait pour un élément observé, prélevé au hasard dans la population, d'avoir la propriété M est un événement E de probabilité inconnue p. De manière perceptive, on sait que lorsqu'on répète cette expérience un grand nombre n de fois, la fréquence f_n de réalisations de E « tend à se stabiliser », et f_n peut être observée aussi proche de p que l'on veut, pourvu que n soit assez grand. Ce phénomène permet de proposer expérimentalement des encadrements possibles de p à partir des valeurs observées des f_n (fourchettes d'échantillonnage) avec d'assez grandes chances de ne pas se tromper.

II. Simulation d'un sondage

En classe de seconde, dans le cadre d'un des thèmes d'études au programme, on peut proposer aux élèves d'exploiter les fonctionnalités du tableur Excel pour mettre en œuvre des simulations de sondages aléatoires simples. Pour cela, conformément à l'exemple qui suit, on utilise les propriétés de la fonction « ALEA » pour simuler le

(*) IREM de Franche-Comté.

prélèvement au hasard d'échantillons de différentes tailles. Le paramètre p qui caractérise le modèle de population considéré peut être introduit dans une cellule cachée et utilisé pour représenter la proportion des éléments de la population qui ont la propriété M. La population est numérotée de 1 à N (dans l'exemple, de 1 à 10^6), et les p_N premiers éléments sont réputés avoir la propriété M. On peut commencer par mettre en évidence la variabilité des fréquences observées suivant les échantillons prélevés (fluctuations d'échantillonnage). Dans l'exemple, on s'intéresse d'abord à des échantillons de taille $n = 100$ puis, en les regroupant, on en considère un de taille 1 000.

La fréquence observée f_n donne une estimation ponctuelle de la proportion p. Bien sûr, on s'intéresse à la précision de cette estimation, c'est-à-dire à un encadrement de « confiance » de p à partir de cette valeur f_n (fourchette de sondage), tel qu'un pourcentage significatif d'échantillons (en principe 95 %) fournisse une fourchette qui contient effectivement p. Le programme de seconde propose une formule « magique » : $\left]f_n - \frac{1}{\sqrt{n}}; f_n + \frac{1}{\sqrt{n}}\right[$, pour déterminer de telles fourchettes, pour des valeurs de p pas trop éloignées de 0,5 (entre 0,3 et 0,7 ; nous verrons comment justifier et étendre ce résultat à toute valeur de p différente de 0 ou 1). Des fourchettes de sondages peuvent alors être obtenues dans de multiples simulations, et les élèves peuvent vérifier qu'environ 95 % d'entre elles contiennent bien la valeur à estimer p.

Le tableau qui suit est un extrait d'une feuille de calcul dans l'exemple proposé.

– La colonne A est obtenue à partir de la fonction « =ALEA() » d'Excel qui donne des chiffres pseudo-aléatoires équirépartis, présentés par groupes de 8 sous la forme de décimales d'un nombre de [0, 1].

– La partie entière du produit de ces nombres par 10^6 permet de simuler le prélèvement au hasard d'éléments dans une population de taille 1 000 000 (colonne B : « =ENT(A1*1000000) »).

– Une valeur pour p ayant été introduite (cellule F14), la comparaison de chacun des nombres précédents avec $p \times 10^6$, traduite en 1 ou 0 (colonne C) par le test « =SI(ENT(100*A1)<100*F14;1;0) », simule pour chaque élément prélevé le fait d'avoir ou non la propriété M, le choix 1 est donc en proportion p dans la population.

– En totalisant par paquets de $n = 100$ la colonne C, on obtient les fréquences observées du 1 dans ces divers échantillons, 10 d'entre elles sont présentées en colonne D (« =SOMME(C1:C100)/100 », puis (C101:C200) etc.).

– Les bornes des fourchettes d'échantillonnage, calculées avec la formule magique (dans ce cas : $]f - 0{,}1 ; f + 0{,}1[$), sont données dans les colonnes F et G, à titre de comparaison.

– Les 10 échantillons étant regroupés, on obtient un échantillon de taille 1 000, avec une fourchette plus étroite (cellules F12 et G12). L'intervalle de confiance théorique

est calculé à partir de la formule indiquée en dessous (cellules F13 et G13), dont la démonstration fait l'objet de la deuxième partie de cet article.

	A	B	C	D	E	F	G
1	0,49716356	497163	0	0,43	Échantillons, taille $n = 100$	0,33	0,53
2	0,85464871	854648	0	0,32		0,22	0,42
3	0,71631158	716311	0	0,36	fourchettes de sondages :	0,26	0,46
4	0,10738629	107386	1	0,31		0,21	0,41
5	0,85212683	852126	0	0,33	$]f - 1/\sqrt{n}\,;\, f + 1/\sqrt{n}[$	0,23	0,43
6	0,94887578	948875	0	0,33		0,23	0,43
7	0,12033745	120337	1	0,38		0,28	0,48
8	0,04542526	45425	1	0,44		0,34	0,54
9	0,33275705	332757	1	0,28		0,18	0,38
10	0,21831516	218315	1	0,33		0,23	0,43
11	0,28453486	284534	1	échantillon de taille 1000, $f =$		0,351	
12	0,62820339	628203	0	fourchette simplifiée :		0,319	0,383
13	0,79683387	796833	0	intervalle de confiance :		0,321	0,381
14	0,91677761	916777	0	Probabilité théorique (cellule cachée) $p =$		0,37	
15	0,88421451	884214	0				
16	0,08202963	82029	1	A = 1000 nombres au hasard équirépartis			
17	0,82429676	824296	0	B = échantillon de 1000 personnes parmi 1 000 000			
18	0,62983241	629832	0	C = préférences de ces 1000 personnes			
19	0,79539564	795395	0	D = fréquences du choix 1 par échantillons de taille 100			
20	0,79191754	791917	0	] F; G [: fourchettes de sondage simplifiées			
21	0,63348029	633480	0	et intervalle de confiance (niveau 0,95) :			
etc. jusqu'à 1 000	…	…		$\left]f - \frac{1{,}96\sqrt{f(1-f)}}{\sqrt{n}}\,;\, f + \frac{1{,}96\sqrt{f(1-f)}}{\sqrt{n}}\right[$			

III. Estimation de p par intervalle de confiance

1) Le modèle probabiliste

Introduisons le modèle probabiliste de cette situation de sondage. Le prélèvement au hasard d'un élément de la population est représenté par une variable aléatoire X_0, variable parente de l'échantillonnage, qui prend la valeur 1 avec probabilité p pour les éléments présentant la propriété M (E est réalisé) et 0 avec probabilité $1 - p$ sinon. La loi de X_0 est donc la loi de Bernoulli B $(1, p)$, d'espérance

$$E(X_0) = p \cdot 1 + (1 - p) \cdot 0 = p$$

et de variance

$$\text{Var}(X_0) = E(X_0^2) - [E(X_0)]^2 = p\,(1 - p).$$

Le prélèvement de l'échantillon des n éléments e_i est représenté par le vecteur aléatoire $X = (X_1, X_2, \ldots, X_n)$, où les X_i sont des répliques successives de X_0. On va faire l'hypothèse de travail que les prélèvements des e_i ne changent pas notablement la proportion p de ceux qui sont de modalité M dans la population. Ceci signifie que la taille de la population est supposée assez grande par rapport à celle de l'échantillon (au moins 100 fois), ou que le prélèvement des e_i est effectué avec remises. Ainsi, la réalisation ou non de l'événement E lors des prélèvements des premiers e_i n'a pas d'effet sur la probabilité de réalisation de E pour les suivants. Pour traduire cela en hypothèse de modèle, on considère que les X_i sont des variables aléatoires indépendantes de même loi de Bernoulli B $(1, p)$ que X_0.

La moyenne $F_n = \frac{1}{n}\sum X_i$ est une variable aléatoire d'espérance $E(F_n) = \frac{1}{n}\sum E(X_i) = p$ et de variance $Var(F_n) = \frac{1}{n^2}\sum Var(X_i) = \frac{p(1-p)}{n}$ (l'indépendance des X_i permet l'additivité des variances).

Sa valeur f_n observée est la fréquence des éléments de modalité M dans l'échantillon (c'est le nombre des X_i prenant la valeur 1 divisé par n). Elle dépend de l'aléa du prélèvement. Cette valeur est prise pour estimer la proportion p inconnue (estimation ponctuelle). Nous verrons que cette démarche intuitive est justifiée par la loi des grands nombres.

L'écart $|f_n - p|$, erreur commise dans l'estimation ponctuelle de p, dépend aussi de l'aléa du prélèvement. On ne peut donc espérer obtenir qu'un contrôle probabiliste *a priori* de $|F_n - p|$, c'est-à-dire limiter le risque de se tromper en donnant un majorant ε de l'erreur que l'on fera si l'on prend pour p la valeur f_n de la fréquence de l'événement E dans un échantillon à prélever. Ce risque est la probabilité que p ne soit pas dans l'intervalle $]f_n - \varepsilon\ ;\ f_n + \varepsilon[$, dit « de confiance », que fournira l'observation de l'échantillon. On désignera par α un majorant de ce risque : $P(|F_n - p| \geq \varepsilon) \leq \alpha$.

2) Le théorème de Bernoulli

Pour évaluer ce majorant α du risque, on a un outil théorique simple, l'inégalité de Bienaymé-Tchebychev, résultat important de la théorie des probabilités :

Soit Y une variable aléatoire de loi quelconque mais dont l'espérance E(Y) et la variance Var(Y) existent. Alors pour tout $\varepsilon > 0$, la probabilité que Y s'écarte de E(Y) de plus que ε est contrôlée par la dispersion de Y. On a :

$$P(|Y - E(Y)| \geq \varepsilon) \leq \frac{Var(Y)}{\varepsilon^2}.$$

Appliquée à la fréquence F_n, cette inégalité donne pour tout $\varepsilon > 0$:

$$P(|F_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2},$$

probabilité qui tend vers 0 quand n tend vers l'infini. On dit que « la fréquence F_n tend vers p en probabilité ». C'est le théorème de Bernoulli, forme la plus simple de la loi (faible) des grands nombres.

En passant à l'événement contraire, cette inégalité s'écrit aussi :

$$P(F_n - \varepsilon < p < F_n + \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

qui montre que pour un ε donné et n assez grand, la probabilité que p soit dans l'intervalle $]F_n - \varepsilon\,;\, F_n + \varepsilon[$ est aussi voisine de 1 que l'on veut.

3) Détermination d'un intervalle de confiance de niveau donné

La valeur $\alpha = \dfrac{p(1-p)}{n\varepsilon^2}$ majore le risque de se tromper en annonçant que p sera dans l'intervalle $]F_n - \varepsilon\,;\, F_n + \varepsilon[$, appelé « intervalle de confiance pour p de niveau $1 - \alpha$ ».

La relation précédente, écrite sous la forme $\varepsilon = \sqrt{\dfrac{p(1-p)}{\alpha}} \cdot \dfrac{1}{\sqrt{n}}$, montre que ε varie comme $\dfrac{1}{\sqrt{n}}$. Le coefficient de proportionnalité $\sqrt{\dfrac{p(1-p)}{\alpha}}$ reflète la « performance » de l'inégalité de Bienaymé-Tchebychev.

On voit aussi que plus le niveau de confiance souhaité sera proche de 1, moins bonne sera la précision de l'estimation proposée.

En majorant grossièrement $p\,(1 - p)$ par 1/4, on obtient une relation simple entre α et la précision ε : $\alpha = \dfrac{1}{4n\varepsilon^2}$. On a ainsi un intervalle de confiance simplifié de niveau $1 - \alpha$, aussi proche de 1 que l'on veut :

$$P\left(F_n - \frac{1}{2\sqrt{\alpha}\sqrt{n}} < p < F_n + \frac{1}{2\sqrt{\alpha}\sqrt{n}}\right) \geq 1 - \alpha.$$

Une fois l'échantillon prélevé, un niveau de confiance étant donné (par exemple 0,95), l'intervalle observé $\left]f_n - \dfrac{1}{2\sqrt{\alpha}\sqrt{n}}\,;\, f_n + \dfrac{1}{2\sqrt{\alpha}\sqrt{n}}\right[$ est une interprétation théorique de la fourchette de sondage, statistiquement obtenue quand on répète un grand nombre de fois cet échantillonnage (pour 95 % des échantillons, on trouve effectivement p dans cette fourchette).

Au niveau de confiance $1 - \alpha$, l'approximation de l'estimation de p par f_n est donc majorée par $\dfrac{1}{2\sqrt{\alpha}\sqrt{n}}$. Pour $\alpha = 0{,}05$, cela donne une demi-fourchette de

$\sqrt{\frac{5}{n}}$, plus de 2,2 fois plus écartée que la demi-fourchette $\frac{1}{\sqrt{n}}$ donnée dans le programme de seconde, laquelle découle d'une amélioration notable de cette méthode de calcul de l'intervalle de confiance basée sur l'inégalité de Bienaymé-Tchebychev. On peut chercher à raffiner cette inégalité, comme nous le verrons dans le paragraphe 5.

Mais, d'une part, la majoration de $p\,(1-p)$ par 1/4 peut paraître trop grossière. D'autre part, l'inégalité théorique de Bienaymé-Tchebychev, ne faisant pas intervenir la loi de la variable d'estimation F_n dans la majoration du risque $P(|F_n - p| \geq \varepsilon)$, est trop générale et ne peut être intéressante dans la pratique : avec $\alpha = 0{,}05$, il faudrait 50 000 observations pour estimer p à 1 % près (avec les mêmes données numériques, le calcul fastidieux publié dans *Ars Conjectandi* en 1713, œuvre magistrale de Jacques Bernoulli, donnerait $n = 68\,100$).

La démarche la plus probabiliste est donc de faire intervenir la loi de F_n. Or $n\,F_n$ suit une loi binomiale B (n, p) dont la mise en œuvre est trop lourde pour n grand. On utilise plutôt l'approximation de cette loi binomiale par une loi normale, résultat majeur dû à Abraham de Moivre (*The Doctrine of Chances*, 1718), et établi entièrement par Pierre Simon Laplace en 1812 dans sa *Théorie analytique des probabilités*.

4) Approximation normale, le théorème de Moivre-Laplace

Le théorème de Moivre-Laplace (forme particulière d'un théorème puissant des probabilités, le « théorème limite central ») permet d'améliorer grandement la performance de l'estimation de p. Ce théorème dit en gros que :

Pour $n > 50$ et pour p pas trop voisin de 0 ou de 1 (ce qui n'est pas trop demander), on fait une erreur négligeable sur la valeur de la probabilité $P(|F_n - p| < \varepsilon)$ en considérant que F_n suit une loi normale $N\left(p ; \sqrt{\frac{p(1-p)}{n}}\right)$, ou encore que la variable $U = \frac{F_n - p}{\sqrt{p(1-p)}}\sqrt{n}$ est normale centrée réduite.

La condition de confiance $P(|F_n - p| < \varepsilon) \geq 1 - \alpha$ s'écrit :

$$P\left(|U| < \varepsilon \frac{\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 1 - \alpha.$$

Si u_α désigne le fractile de cette loi vérifiant $P(|U| < u_\alpha) = 1 - \alpha$, on a $u_\alpha = \varepsilon \frac{\sqrt{n}}{\sqrt{p(1-p)}}$, et on obtient l'intervalle de confiance pour p au niveau $1 - \alpha$:

$$\left]F_n - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; F_n + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right[.$$

Remarquons que la relation qui lie ε, n et α ressemble à celle que nous avions obtenue en appliquant l'inégalité de Bienaymé-Tchebychev : $\varepsilon = u_\alpha \sqrt{p(1-p)} \cdot \frac{1}{\sqrt{n}}$. À titre de comparaison, pour α = 0,05, on a $u_\alpha \cong 1{,}96$, alors que l'on avait $\frac{1}{\sqrt{\alpha}} \cong 4{,}47$.

Mais, dans notre situation de sondages, p est inconnu. La variance $\frac{p(1-p)}{n}$ de la loi de F_n peut être obtenue en y estimant ponctuellement p par la fréquence f_n de l'événement E observée dans l'échantillon prélevé. On fait alors une erreur négligeable (au second ordre près quand n est grand) dans le calcul des probabilités liées à F_n. L'intervalle de confiance pour p au niveau $1 - \alpha$ donne alors la fourchette théorique :

$$\left]f_n - u_\alpha \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}} ; f_n + u_\alpha \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}}\right[.$$

Avec $\alpha = 0{,}05$, $u_\alpha \cong 1{,}96$ est très voisin de 2. En majorant comme précédemment $f_n(1-f_n)$ par 1/4 quand f_n n'est pas trop proche de 0 ou 1, on perd un peu en précision mais on simplifie notablement l'expression du demi-écart $\varepsilon = u_\alpha \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}}$ de la fourchette de sondage théorique qui est alors majoré par $\frac{1}{\sqrt{n}}$. Au niveau de confiance $1 - \alpha = 0{,}95$, on retrouve donc l'intervalle proposé dans le thème d'étude du programme de seconde :

$$\left]f_n - \frac{1}{\sqrt{n}} ; f_n + \frac{1}{\sqrt{n}}\right[.$$

Dans ces conditions, il suffit d'un échantillon de taille n = 10 000 (c'est encore cher) pour estimer p à 1 % près, avec une probabilité 0,95 de ne pas se tromper. Par contre, si on accepte une estimation de p à 3 % près, il suffit que $n \geq 1\,112$ $\left(\frac{1}{\sqrt{n}} = 0{,}03\right)$, taille approximative des sondages les plus courants.

Le tableau suivant donne une idée des précisions ε, demi-écarts des fourchettes théoriques exprimés en pourcentages, obtenues pour différentes tailles d'échantillons et différents niveaux de confiance pour estimer une proportion.

$1-\alpha$ \ n / u_a		500	800	1 000	2 000	10 000
0,9	1,65	3,7	2,9	2,6	1,8	0,8
0,95	1,96	4,4	3,5	3,1	2,2	1
0,99	2,60	5,8	4,6	4,1	2,9	1,3

On voit donc qu'avec des échantillons de taille 1 000, le hasard des prélèvements peut à lui seul justifier une différence de 3 % dans les estimations affichées pour une proportion inconnue p, avec d'ailleurs 5 chances sur 100 pour que l'écart entre p et une valeur annoncée soit réellement supérieur à 3 %. Il est donc abusif de donner p à 1 % près comme on le rencontre souvent, et d'autant plus abusif de disserter sur une variation de moins de 3 % entre des valeurs estimées.

5) Une inégalité optimale pour contrôler le risque α

Pour les grandes valeurs de n, on peut nettement améliorer l'inégalité de Bienaymé-Tchebychev dont la démonstration procède d'une majoration trop brutale. Une inégalité plus fine a été établie en 1924 par Serge Bernstein(1). Elle montre que la décroissance du risque est exponentielle. Notons $q = 1 - p$ et considérons un écart ε tel que $0 < \varepsilon < \inf(p, q)$. Posons

$$r(p,\varepsilon) = (p+\varepsilon)\ln\left(1+\frac{\varepsilon}{p}\right) + (q-\varepsilon)\ln\left(1-\frac{\varepsilon}{q}\right).$$

On montre que $r(p, \varepsilon) > 0$. Alors, pour tout $n \geq 1$, on a :

$$\mathrm{P}(|\mathrm{F}_n - p| \geq \varepsilon) \leq \mathrm{e}^{-n\cdot r(p,\varepsilon)} + \mathrm{e}^{-n\cdot r(q,\varepsilon)}.$$

On démontre que cette inégalité est optimale. On ne peut donc obtenir une meilleure majoration du risque sans faire intervenir la loi de F_n.

Quand ε est petit devant p et q, on peut obtenir une relation asymptotique entre un niveau de confiance $1 - \alpha$ donné, la précision ε et la taille n de l'échantillon prélevé. En effet, en posant $r = \inf[r(p,\varepsilon) ; r(q,\varepsilon)]$, on majore simplement le second membre de cette inégalité :

$$\mathrm{P}(|\mathrm{F}_n - p| \geq \varepsilon) \leq \mathrm{e}^{-nr}(1+\mathrm{e}^{-n|r(p,\varepsilon)-r(q,\varepsilon)|}) \leq 2\,\mathrm{e}^{-nr}.$$

La valeur $\alpha = 2\mathrm{e}^{-nr}$ majore donc le risque que p soit en dehors de l'intervalle de confiance, condition que l'on peut écrire $n \cdot r = \ln(2/\alpha)$ pour mieux faire apparaître la relation entre la taille n de l'échantillon et la précision ε de l'estimation, nous pourrons ainsi cerner la performance de cette inégalité de Bernstein.

(1) Pour sa démonstration, très accessible, on pourra se reporter au livre d'Emmanuel Lesigne : *Pile ou Face, une introduction aux théorèmes limites du Calcul des Probabilités*, paru chez Ellipses en 2001, ouvrage que Paul-Louis Hennequin m'a aimablement signalé.

Comme il se doit, ε tend vers 0 quand n tend vers l'infini. Dans les deux cas précédents, nous avons effectivement vu que pour une proportion p et un niveau de confiance $1 - \alpha$ fixés, le produit $n \cdot \varepsilon^2$ reste constant quand la taille de l'échantillon varie ($n \cdot \varepsilon^2 = \dfrac{pq}{\alpha}$ avec l'inégalité de Bienaymé-Tchebychev et $n \cdot \varepsilon^2 = u_\alpha^2 pq$ par le théorème de Moivre-Laplace). Quand ε tend vers 0, les fonctions $r(p,\varepsilon)$ et $r(q,\varepsilon)$ sont équivalentes à $\dfrac{\varepsilon^2}{2pq}$ (on le voit en développant les logarithmes figurant dans $r(p,\varepsilon)$ et $r(q,\varepsilon)$ à l'ordre 2 en ε), on obtient donc $n \cdot r \cong \dfrac{n\varepsilon^2}{2pq}$, d'où

$$\varepsilon \cong \sqrt{2\ln\left(\frac{2}{\alpha}\right)} \cdot \sqrt{p(1-p)} \cdot \frac{1}{\sqrt{n}}.$$

Dans cette estimation, comme dans les deux précédentes, la précision ε est asymptotiquement proportionnelle à $\sqrt{p(1-p)} \cdot \dfrac{1}{\sqrt{n}}$, avec comme coefficient de proportionnalité $\sqrt{2\ln\left(\dfrac{2}{\alpha}\right)}$. À titre de comparaison, avec $\alpha = 0{,}05$ et en majorant $\sqrt{p(1-p)}$ par 1/2, on obtient pour majorer ε la valeur $\dfrac{1{,}36}{\sqrt{n}}$, intermédiaire entre $\dfrac{2{,}24}{\sqrt{n}}$ donné par l'inégalité de Bienaymé-Tchebychev et $\dfrac{1}{\sqrt{n}}$ obtenu par le théorème de Moivre-Laplace.

Pour comparer plus précisément les performances de ces différentes estimations, prenons $p = q = 1/2$ (le cas le plus défavorable) et un niveau de confiance de 0,95. Le tableau suivant donne pour chacune d'elles et pour différentes tailles d'échantillons les demi-fourchettes ε obtenues, exprimées en pourcentages.

n	500	800	1 000	2 000	10 000
Bienaymé-Tchebychev	10	7,9	7,1	5	2,2
Moivre-Laplace	4,4	3,5	3,1	2,2	1
S. Bernstein	6,1	4,8	4,3	3	1,4