Statistiques et Probabilités

académie Bordeaux

MINISTÈRE DE L'ÉDUCATION NATIONALE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

RÉPUBLIQUE FRANÇAISE

Équipe Académique Mathématiques - 2009

Les statistiques au collège

Exercice 2 **4 points**

Dans un collège, une enquête a été menée sur « le poids des cartables des élèves ».
Pour cela, on a pesé le cartable de 48 élèves du collège.
Les résultats de cette enquête sont inscrits dans le tableau ci dessous :

Poids en kg	1	2	3	4	5	6	7	8	9	10
Effectif	1	2	4	2	5	11	8	8	3	4

1. Calculer l'étendue de cette série statistique.
2. Déterminer la médiane de cette série statistique.
3. Déterminer, les valeurs du premier quartile et du troisième quartile de la série.
4. Une personne affirme :
   « Plus des trois quarts des 48 élèves viennent en cours avec un cartable qui pèse 5 kg ou plus ».
   A t-elle raison ? Justifier votre réponse.

DNB 2009 -Asie

Les statistiques en seconde

Résumé des notions abordées au collège

Les notions de moyenne, médiane, étendue, quartiles ont été développées au collège ainsi que leurs interprétations.
Ces notions pourront être sollicitées en classe de seconde en lien avec les autres disciplines.
(par ex : mesures expérimentales en sciences physiques).

Analyse de données

La classe de seconde est l'occasion d'une part de consolider l'utilisation des fonctions statistiques des calculatrices d'autre part de traiter, à l'aide du tableur, des séries statistiques comportant un grand nombre de données en lien avec des situations réelles

http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/france-departements.asp

http://www.ined.fr/fr/pop_chiffres/france/mortalite_causes_deces/esperance_vie/

Les probabilités au collège

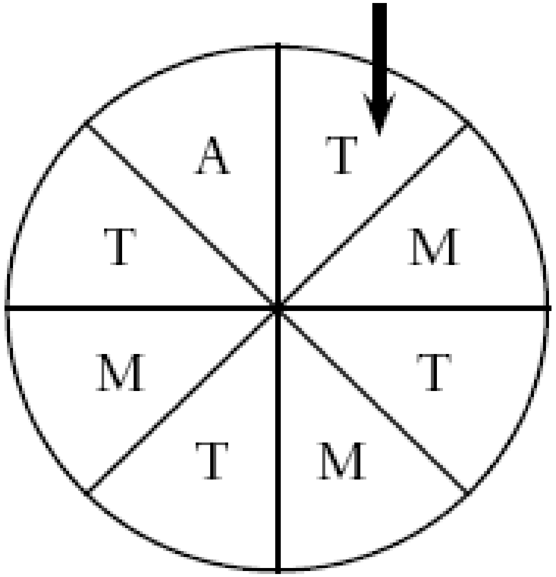
Trois personnes, Aline, Bernard et Claude ont chacune un sac contenant des billes.
Chacune tire au hasard une bille de son sac.
1. Le contenu des sacs est le suivant :
Sac d'Aline : 5 billes rouges
Sac de Bernard : 10 billes rouges et 30 billes noires
Sac de Claude : 100 billes rouges et 3 billes noires
Laquelle de ces personnes a la probabilité la plus grande de tirer une bille rouge ?
2. On souhaite qu'Aline ait la même probabilité que Bernard de tirer une bille rouge.
Avant le tirage, combien de billes noires faut-il ajouter pour cela dans le sac d'Aline ?

DNB 2009 -Métropole

Les probabilités au collège

A un stand du «Heiva », on fait tourner la roue de loterie ci-dessous.
On admet que chaque secteur a autant de chance d'être désigné.
On regarde la lettre désignée par la flèche : A, T ou M, et on considère les évènements suivants :

- A : « on gagne un autocollant» ;
- T : « on gagne un tee-shirt » ;
- M: « on gagne un tour de manège ».



1. Quelle est la probabilité de l'évènement A ?
2. Quelle est la probabilité de l'évènement T?
3. Quelle est la probabilité de l'évènement M?
4. Exprimer à l'aide d'une phrase ce qu'est l'évènement non A puis donner sa probabilité.

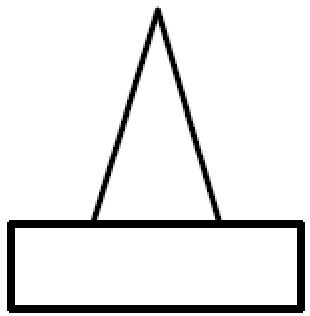
DNB 2009 -Polynésie

Les probabilités au collège

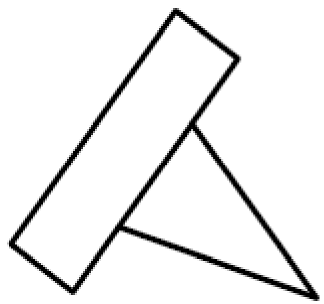
1.4. Notion de probabilité [Thèmes de convergence]	- Comprendre et utiliser des notions élémentaires de probabilité. - Calculer des probabilités dans des contextes familiers.	La notion de probabilité est abordée à partir d'expérimentations qui permettent d'observer les fréquences des issues dans des situations familières (pièces de monnaie, dés, roues de loteries, urnes, etc.). La notion de probabilité est utilisée pour modéliser des situations simples de la vie courante. Les situations étudiées concernent les expériences aléatoires à une ou *à deux épreuves.*

Les probabilités au collège

L'approche fréquentiste nécessite que des fréquences soient observées expérimentalement ; le lancer d'une punaise (pouvant tomber suivant la position « 1 » ou la position « 0 » ci-dessous) a longtemps servi d'exemple dans les pays anglo-saxons :



1



0

L'intérêt du lancer de punaise réside dans le fait que seule l'expérimentation permet de proposer une probabilité au résultat « 1 ».

Les probabilités au collège

- **Une approche fréquentiste.**
- **Réalité**

➔ Expérimentation

➔ Initiation au calcul des probabilités

Un objectif final (dont nous ne savons pas s'il peut être atteint à ce stade) serait : savoir faire la différence de nature entre une fréquence observée (a posteriori) qui est du domaine de la statistique, et une probabilité (déterminée a priori).

Les probabilités au lycée

Objectifs visés par l'enseignement des statistiques et probabilités à l'occasion de résolutions de problèmes

dans le cadre des probabilités, **rendre les élèves capables :**

• d'étudier et modéliser des expériences relevant de l'équiprobabilité

• de proposer un modèle probabiliste à partir de l'observation de fréquences

• d'interpréter des événements de manière ensembliste

• de mener à bien des calculs de probabilité.

Les situations étudiées concernent des expériences à une ou plusieurs épreuves.

A partir d’une situation concrète

➔ **Simulation**

➔ **Fluctuation des échantillons**

La répétition d’expériences aléatoires peut donner lieu à l’écriture d’algorithmes

➔ **Convergence vers les probabilités**

Les distributions de probabilité peuvent être estimées par observation de la stabilisation des fréquences sur de longues séries d’expériences

➔ **Mise en place d’un modèle**

Les distributions de probabilité peuvent être estimées par des considérations géométriques ou physiques en référence à l’équiprobabilité

Les probabilités au lycée

travail sur le modèle

Le langage des évènements

- **Une distribution de probabilité sur un ensemble fini Ω est définie par la donnée des probabilités des éléments de Ω.**
- **Un événement est défini comme un sous-ensemble de Ω.**

Les représentations

- **Tableaux croisés**
- **Arbres des possibles**
- **Arbres pondérés**

  (*toute connaissance sur le conditionnement est hors-programme*)

Les probabilités au lycée

À propos des arbres pondérés

Exemple (tiré de Bulletin de l'APMEP)

Un vote a eu lieu dans une ville. Nous disposons des informations suivantes :

Scrutin

Groupe I : électeurs de moins de 35 ans ; 38% de l'ensemble des électeurs.

Groupe II : électeurs de 35 à 60 ans ; 43% de l'ensemble des électeurs.

Groupe III : électeurs de plus de 60 ans ; 19% de l'ensemble des électeurs.

Taux de participation

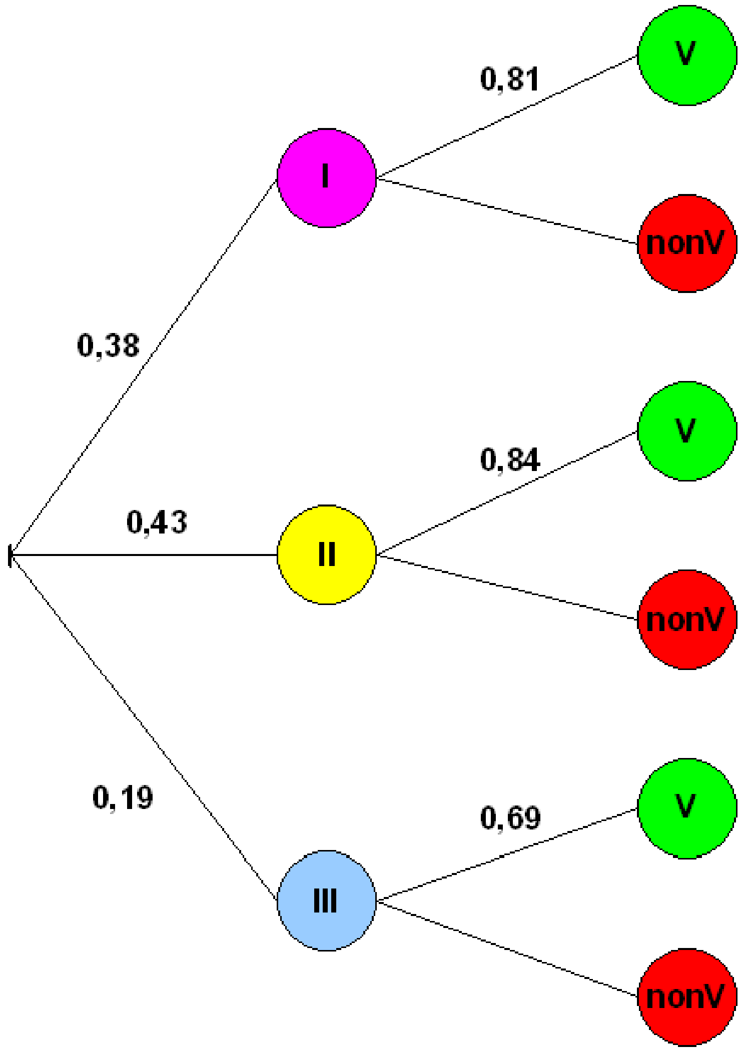
Groupe I : 81%

Groupe II : 84%

Groupe III : 69%

- Représenter ces informations sous forme d'un arbre pondéré.
- Utiliser cet arbre pour répondre aux questions suivantes :
- On choisit un électeur au hasard. Quelle est la probabilité qu'il ait voté ?
- Quel est le taux de participation au scrutin ?

Les probabilités au lycée



Un calcul de pourcentages conduit à admettre que, de manière générale, la probabilité “d’un chemin” est égale au produit des probabilités “rencontrées le long de ce chemin”.

Pour trouver la probabilité demandée, il suffit d’additionner les probabilités des chemins qui aboutissent à V.

Les probabilités au lycée

Le programme de seconde doit :

- **initier au modèle probabiliste**
- **amener à réfléchir à l'aide de ce modèle.**

La statistique inférentielle au lycée

« Former les élèves en statistique, c'est leur donner les moyens de développer une forme de pensée critique sans laquelle ils seront exclus du débat politique et social. »

Dans le programme 2009

- **Fluctuation des échantillons**
- **Estimation par « fourchette de sondage »**
- **Prise de décision**

Fluctuation des échantillons

Considérons une urne « de Bernoulli »
(la population) contenant une proportion *p* de boules blanches, dont on extrait *n* boules ;
la proportion de boules blanches dans le tirage (ou échantillon) est notée *f*.

p est connu

On note X la variable aléatoire correspondant au nombre de boules blanches dans un échantillon de taille n.

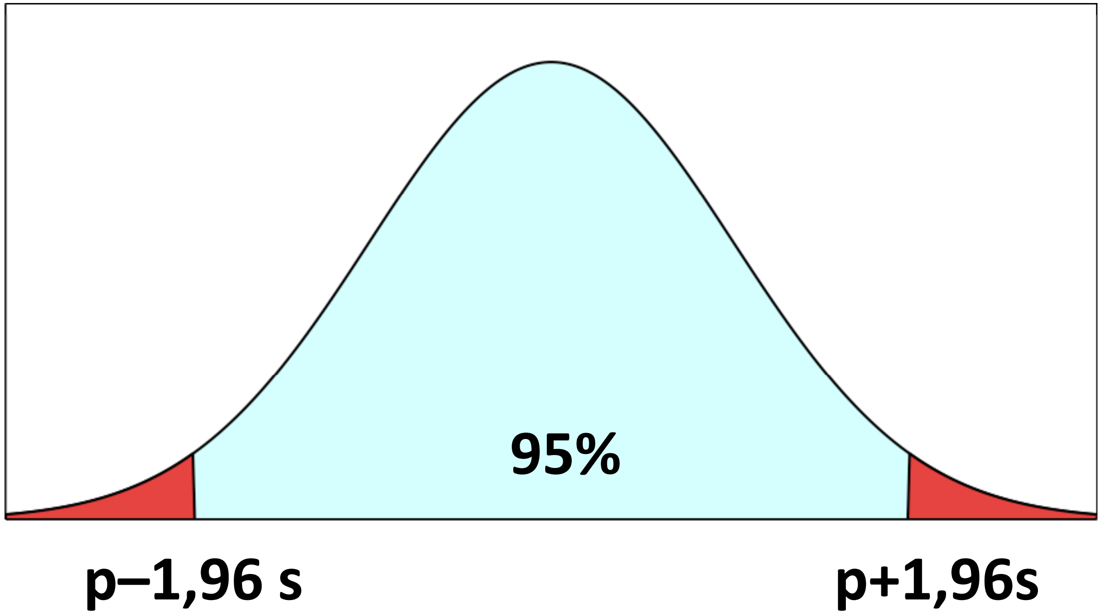
X suit la loi binomiale de paramètres n et p.

Un raisonnement probabiliste, permet d'approcher la variable aléatoire $f = X/n$

par la loi normale $\mathcal{N}\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$.

Sous certaines conditions sur n et p :

$n \geq 25$ et p compris entre 0,2 et 0,8.



Si T suit la loi normale d'espérance p et d'écart type s , on a : $P(p - 1{,}96s \leq T \leq p + 1{,}96s) \approx 0{,}95$.

p est connu

La probabilité que *f* appartienne à l'intervalle

$$\left[p-\frac{1{,}96\sqrt{p(1-p)}}{\sqrt{n}};p+\frac{1{,}96\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

est environ égale à 0,95.

Cet intervalle est l'intervalle de fluctuation au seuil de 95%

Remarquons que le produit $p(1-p)$ est toujours inférieur à 1/4. On élargit donc l'intervalle de fluctuation au seuil de 95% à :

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

Les situations étudiées correspondent aux cas où les nombres n et p vérifient $n \geq 25$

et p compris entre 0,2 et 0,8.

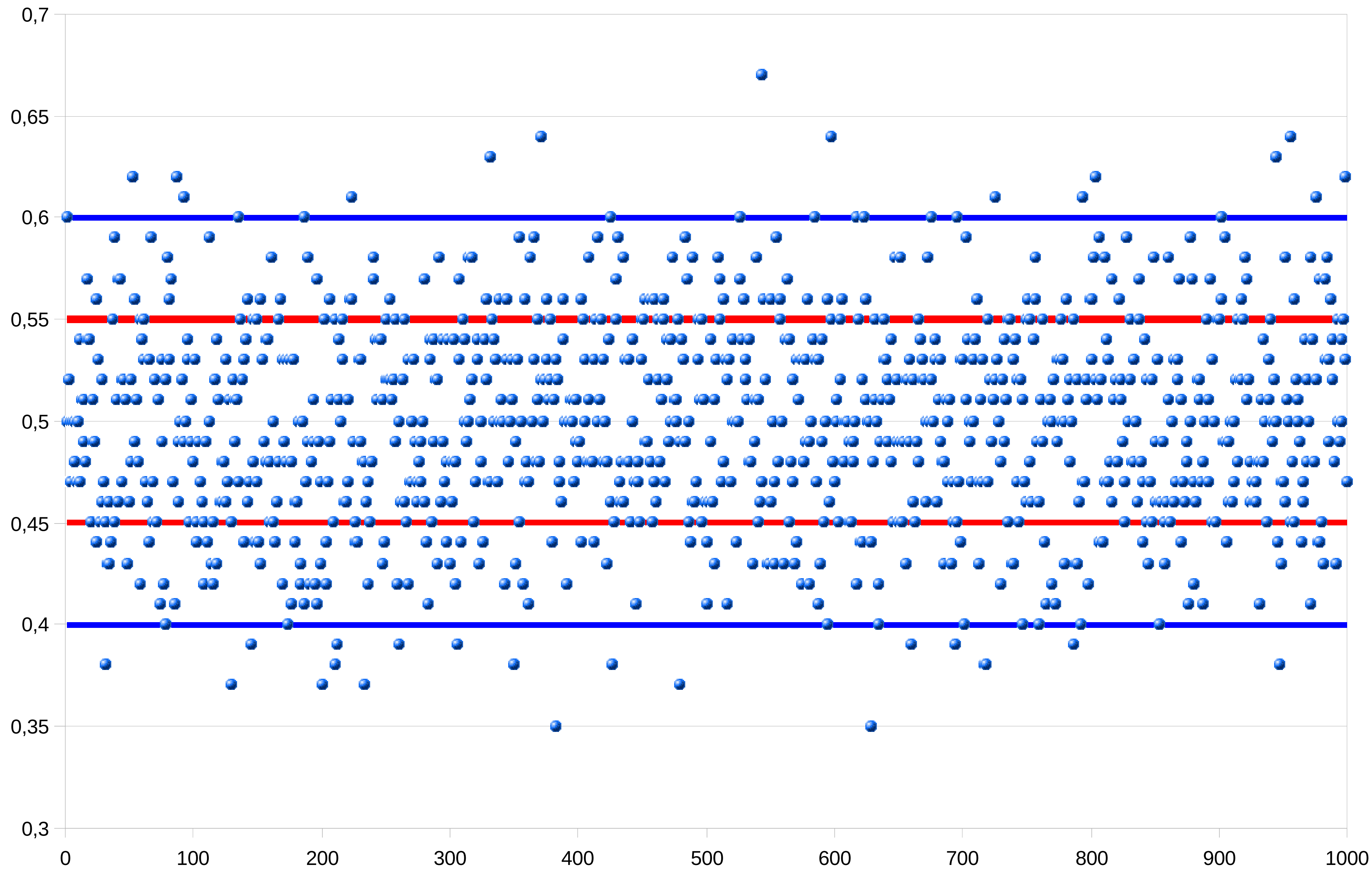
La connaissance de ces conditions n'est pas exigible. La formule de l'intervalle est donnée.

On effectue 1000 tirages de taille $N = 100$ pour $p = 0,5$.

Dans le graphique ci-après, chaque point représente la fréquence f pour un des 1000 sondages de taille 100. Des intervalles de fluctuation ont été représentées.

En bleu: l'intervalle au seuil de 95%

En rouge: l'intervalle au seuil de 68%



0,7
0,65
0,6
0,55
0,5
0,45
0,4
0,35
0,3
0
100
200
300
400
500
600
700
800
900
1000

Estimation
ou
« fourchette de sondage »

Si p est inconnu mais que l'on procède à un tirage donnant une valeur de f ; on peut dire que dans environ 95% des tirages on a :

$$p \in \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$$

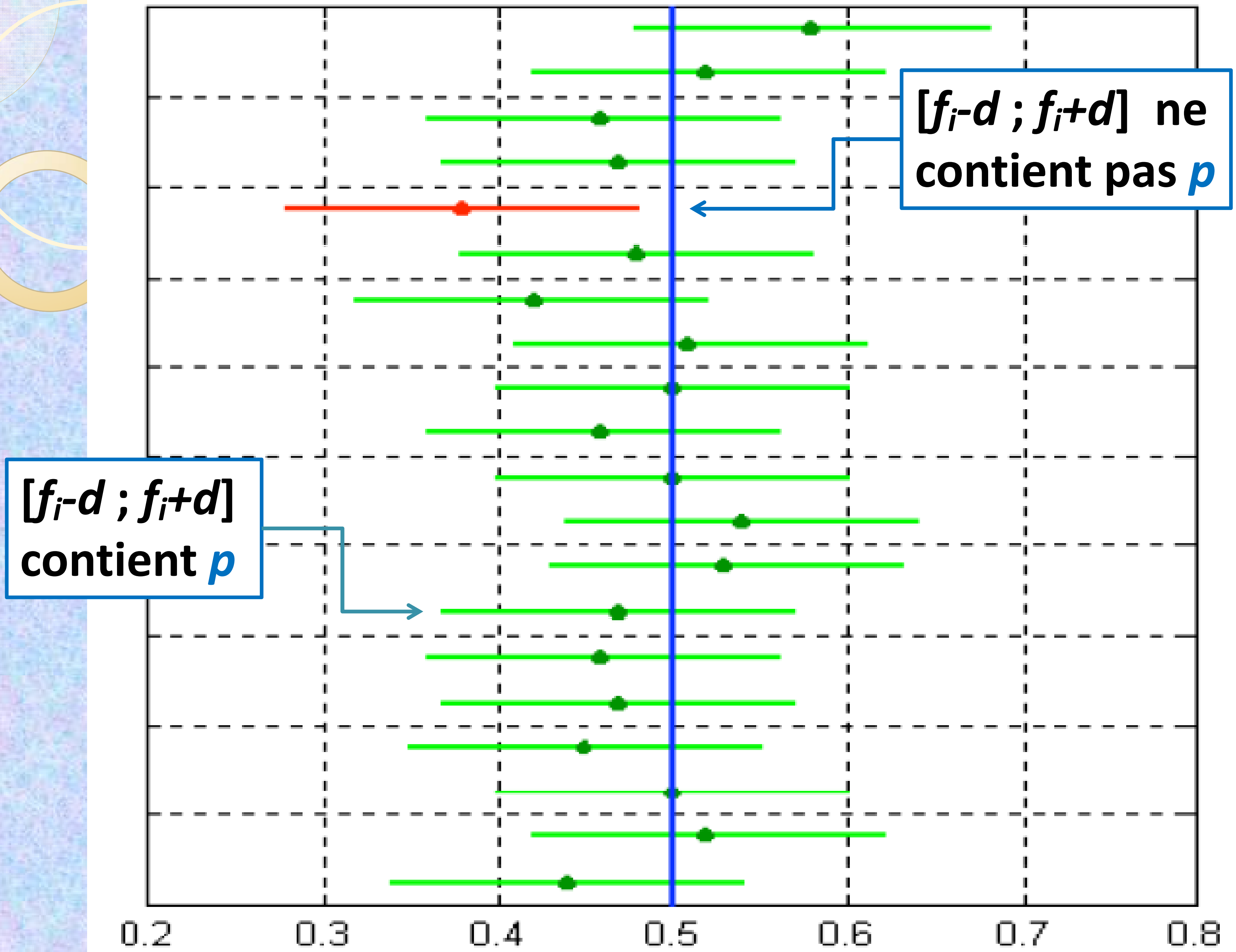
Cet intervalle est appelé intervalle de confiance ou fourchette de sondage.

Si on calcule l'intervalle de confiance précédent pour 100 sondages indépendants de taille *n* sur la même urne, on peut s'attendre à ce qu'environ :

95 contiennent la vraie valeur de *p*

5 ne la contiennent pas.

Ceci est illustré dans la figure suivante.



[f_i-d ; f_i+d] ne contient pas p
[f_i-d ; f_i+d] contient p
0.2
0.3
0.4
0.5
0.6
0.7
0.8

On ne connaît qu'une seule valeur de f (c'est-à-dire pas grand chose) et c'est un raisonnement de nature inductive qui permet d'estimer p à l'aide de l'intervalle de confiance.

Avant le tirage, il y avait 95 chances sur 100 d'avoir un intervalle contenant p.

Après le tirage p est ou n'est pas dans l'intervalle de confiance.

Prise de décision

Un médecin de la santé publique s'interroge sur la proportion de patients souffrant d'hypertension dans sa commune.
Une récente étude dans des populations semblables indique une proportion égale à 17%.
Il étudie un échantillon de taille 1000 et calcule la fréquence f obtenue.

f appartient-elle à l'intervalle de fluctuation [0,14 ; 0,20] au seuil de 95% ?

Si la réponse est non

il rejette l'hypothèse $p = 0{,}17$ avec un risque d'erreur de 5%.

Si la réponse est oui

l'hypothèse est acceptable sans connaître le risque d'erreur.