Intervalle de fluctuation
Intervalle de confiance

Lequel utiliser ?

Pourquoi ?

Comment simuler ?

Animation nouveaux programmes de mathématiques
Terminale STI2D – Académie de Créteil, jeudi 3 mai 2012

Deux intervalles pour deux problématiques

- **Intervalle de fluctuation = prise de décision**

Population

Proportion *p* connue ou supposée

Échantillon de taille *n* prélevé au hasard

Fréquence *f* ?

Population

Proportion *p*?

Échantillon de taille *n* prélevé au hasard

Fréquence *f* connue

Sur les trois années de lycée

	Intervalle de fluctuation au seuil de 95 % (*p* connue)	Intervalle de confiance au niveau de confiance 95% (*f* connue)
2nde	**Conditions:** $n \geq 25$ et $0{,}2 \leq p \leq 0{,}8$ **Forme:** $\left[p-\frac{1}{\sqrt{n}}; p+\frac{1}{\sqrt{n}}\right]$	**Sensibilisation**
1ère	**Conditions: aucune.** **Forme: obtenue à l'aide de la loi binomiale.**	
Term STI2D	**Conditions:** $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$. **Forme:** $\left[p-1{,}96\sqrt{\frac{p(1-p)}{n}}; p+1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$	**Conditions:** $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$. **Forme:** $\left[f-1{,}96\sqrt{\frac{f(1-f)}{n}}; f+1{,}96\sqrt{\frac{f(1-f)}{n}}\right]$

Utilisations de l’intervalle de fluctuation

- **Programme:**

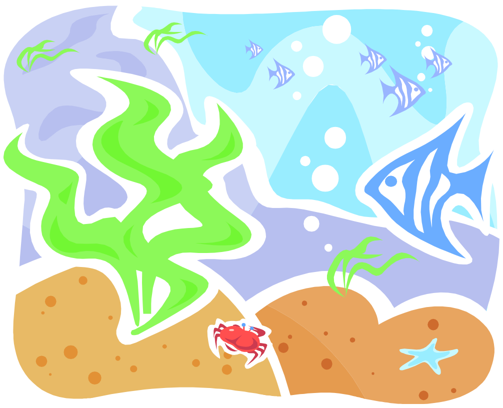
Prise de décision et estimation Intervalle de fluctuation d’une fréquence.	• Connaître l’intervalle de fluctuation asymptotique à 95 % d’une fréquence obtenue sur un échantillon de taille n : $\left[p-1{,}96\sqrt{\frac{p(1-p)}{n}}\ ,\ p+1{,}96\sqrt{\frac{p(1-p)}{n}}\right]$ lorsque la proportion p dans la population est connue. • Exploiter un tel intervalle de fluctuation pour rejeter ou non une hypothèse sur une proportion.	On fait observer que cet intervalle est proche de celui déterminé en première à l’aide de la loi binomiale, dès que $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

L’utilisation classique de l’intervalle de fluctuation: rejet ou non d’une hypothèse sur une proportion

- **Contexte:**

Dans un pays lointain, 10% des plages étaient atteintes par des algues toxiques. On a modifié le processus de rejets chimiques: on admet que le nouveau processus de rejet, très différent du précédent, pourrait augmenter ou diminuer cette fréquence.

On veut établir, puis mettre en œuvre, une procédure statistique permettant de décider, au seuil de 5%, si le nouveau processus de rejets a, ou non, un impact significatif, dans un sens ou dans un autre, sur la fréquence d'apparition de ces algues.



- **Question 1:**

énoncer une règle de décision permettant de rejeter, ou non, au seuil de décision de 5%, l'hypothèse selon laquelle 10% des plages sont touchées par ce type d'algues après la mise en place du nouveau procédé, à l'aide d'un échantillon aléatoire de 150 plages.

- **Réponse:**

* on fait l’hypothèse que la proportion de plages polluées après la mise en place du procédé est p=0,1;
* on prend un échantillon de n=150 plages et on en détermine la fréquence f de plages polluées.
* on détermine l’intervalle de fluctuation I au seuil de 95% correspondant à n et p.
* on prend une décision:

- soit f n’est pas dans I alors on rejette l’hypothèse de 10% de plages polluées après la mise en place du procédé, avec un risque de se tromper dans 5 % des cas;
- soit f est dans I alors on ne rejette pas l’hypothèse de 10% de plages polluées.

- **Question 2:**

Sur un échantillon aléatoire de 150 plages, on constate que 9 plages sont atteintes. Peut-on, au seuil de décision de 5%, rejeter l'hypothèse précédente de 10% de plages polluées ?

- **Réponse:**

* Ici: p=0,1 et n=150 avec $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$.

* On détermine l'intervalle $I = \left[p - 1{,}96\sqrt{\frac{p(1-p)}{n}} ; p + 1{,}96\sqrt{\frac{p(1-p)}{n}} \right]$: [0,052 ; 0,148].

* On constate que f=9/150=0,06 appartient à [0,052 ; 0,148].

* On en déduit qu'au seuil de décision de 5%, on ne rejette pas l'hypothèse de 10% de plages polluées: le hasard seul peut expliquer l'écart entre f =0,06 et p=0,1 !

Une autre utilisation de l'intervalle de fluctuation: l'homogénéité de lots dans une production

- **Contexte:** la proportion d'ampoules à économie d'énergie non-conformes dans la production d'une entreprise est p=0,07. L'entreprise souhaite fournir des lots d'ampoules pour lesquels elle puisse « garantir » qu'environ 95% d'entre eux ont une fréquence d'ampoules non-conformes entre 0,06 et 0,08.

- **Question:** quelle taille minimale n du lot à prendre pour répondre à cette contrainte?

- **Traduction:** chercher le plus petit entier naturel n tel que $L_n \leq 0{,}02$ où L_n est la longueur de l'intervalle de fluctuation au seuil de 95% soit, avec p=0,07:

$$2\times 1{,}96\sqrt{\frac{0{,}0651}{n}} \leq 0{,}02$$

- **Résolution** de l'inéquation : algébrique ou algorithmique: $n \geq 2501$ (travail sur la notion de seuil).

- **Conclusion:** conditionnement minimal par lots de 2501 ampoules.

Utilisations de l'intervalle de confiance

- **Programme:**

Contenus	Capacités attendues	Commentaires
Intervalle de confiance d'une proportion.	• Estimer une proportion inconnue avec un niveau de confiance de 95 % par l'intervalle : $\left[f - 1{,}96\sqrt{\frac{f(1-f)}{n}}\,,\ f + 1{,}96\sqrt{\frac{f(1-f)}{n}}\right]$ calculé à partir d'une fréquence f obtenue sur un échantillon de taille n.	Cette expression de l'intervalle de confiance, pour n assez grand, est admise. On constate par simulation que, pour $n \geq 30$, sur un grand nombre d'intervalles de confiance, environ 95 % contiennent la proportion à estimer.
	• Juger de l'égalité de deux proportions à l'aide des intervalles de confiance à 95 % correspondant aux fréquences de deux échantillons de taille n.	La différence entre les deux fréquences observées est considérée comme significative quand les intervalles de confiance à 95 % sont disjoints. C'est l'occasion d'étudier des méthodes statistiques pratiquées dans les disciplines scientifiques ou technologiques.

Une première utilisation: l’estimation d’une proportion inconnue

- **Contexte:**

Un industriel fabrique des smartphones. Pour contrôler la qualité de la production, il en teste 200: 92% fonctionnent correctement.

- **Question:**

Qu’en conclure sur l’ensemble de la production?

- Ici: n=200 et f=0,92.

On détermine l'intervalle $\left[f-1{,}96\sqrt{\frac{f(1-f)}{n}}\,;\,f+1{,}96\sqrt{\frac{f(1-f)}{n}}\right]$ de confiance au niveau de confiance de 95%: [0,882 ; 0,958].

- On conclut que la probabilité que p soit dans [0,882 ; 0,958] est au moins égale à 0,95:

NON!

p est un réel fixe qui est ou qui n'est pas dans [0,882 ; 0,958]: il n'y a pas d'aléatoire ici.

On conclut qu'à partir de cet échantillon de 200 smartphones, la proportion de smartphones fonctionnant correctement sur l'ensemble de la production est dans l'intervalle [0,882 ; 0,958] au niveau de confiance de 95%.

Une deuxième utilisation: la notion de différence significative

- **Contexte:**

Durant 6 mois, l’entreprise de smartphones travaille à améliorer la qualité de sa production. Un nouvel échantillon de 200 smartphones est prélevé: 97% fonctionnent correctement.

- **Question:**

Est-il raisonnable, au niveau de confiance de 95%, de penser que la production s’est améliorée?

- **Avant « amélioration » :**
  - n=200 et f=0,92.
  - Intervalle de confiance: [0,882 ; 0,958].

- **Après « amélioration » :**
  - n=200 et f=0,97.
  - Intervalle de confiance: [0,946 ; 0,994].

- **Conclusion:**

$[0{,}882\ ;\ 0{,}958] \cap [0{,}946\ ;\ 0{,}994] \neq \varnothing$

donc la différence entre les deux échantillons n'est pas significative: il n'y a pas d'évolution significative de la qualité de la production au niveau de confiance de 95%.

Un exemple de simulation

- **Programme:**

Contenus	Capacités attendues	Commentaires
Intervalle de confiance d'une proportion.	• Estimer une proportion inconnue avec un niveau de confiance de 95 % par l'intervalle : $\left[f-1{,}96\sqrt{\frac{f(1-f)}{n}}, f+1{,}96\sqrt{\frac{f(1-f)}{n}}\right]$ calculé à partir d'une fréquence f obtenue sur un échantillon de taille n.	Cette expression de l'intervalle de confiance, pour n assez grand, est admise. On constate par simulation que, pour $n \geq 30$, sur un grand nombre d'intervalles de confiance, environ 95 % contiennent la proportion à estimer.

- **Pourquoi simuler?** Donner du sens à la notion d'intervalle de confiance avec une visualisation graphique.

- **Contexte:**
- p: proportion inconnue d’un caractère C d’une population.
- constitution de 100 échantillons aléatoires de taille n=1000.

- **Question :**
  Comment construire et exploiter une simulation à partir de la fréquence f du caractère C pour chacun des 100 échantillons?

Présentation de la feuille de calcul

	A	B	C	D	E
1	valeur de *p*	[redacted]			
2	nombres d'échantillons où *p* appartient à l'intervalle				
3					
4	appartenance de *p* à l'intervalle				
5	borne inférieure de l'intervalle				
6	borne supérieure de l'intervalle				
7	fréquence sur l'échantillon				
8	échantillon n°	n°1	n°2	n°3	n°4
9	1				
10	2				
11	3				
12	4				
13	5				
14	6				

Valeur de *p* cachée

Formule: =ENT(ALEA()+B1)
(à dupliquer dans la plage B9:CW1008).

On obtient 1 avec la probabilité *p* et 0 avec la probabilité 1–*p*.

	A	B	C	D	E
1	valeur de *p*	[redacted]			
2	nombres d'échantillons où *p* appartient à l'intervalle	96			
3					
4	appartenance de *p* à l'intervalle	OUI	OUI	OUI	OUI
5	borne inférieure de l'intervalle	0,815	0,823	0,830	0,846
6	borne supérieure de l'intervalle	0,861	0,867	0,874	0,888
7	fréquence sur l'échantillon	0,838	0,845	0,852	0,867
8	échantillon n°	n°1	n°2	n°3	n°4
9	1	1	1	0	1
10	2	1	0	1	1
11	3	1	1	1	1
12	4	1	1	1	1
13	5	1	1	1	1
14	6	1	1	1	1
15	7	1	1	0	1

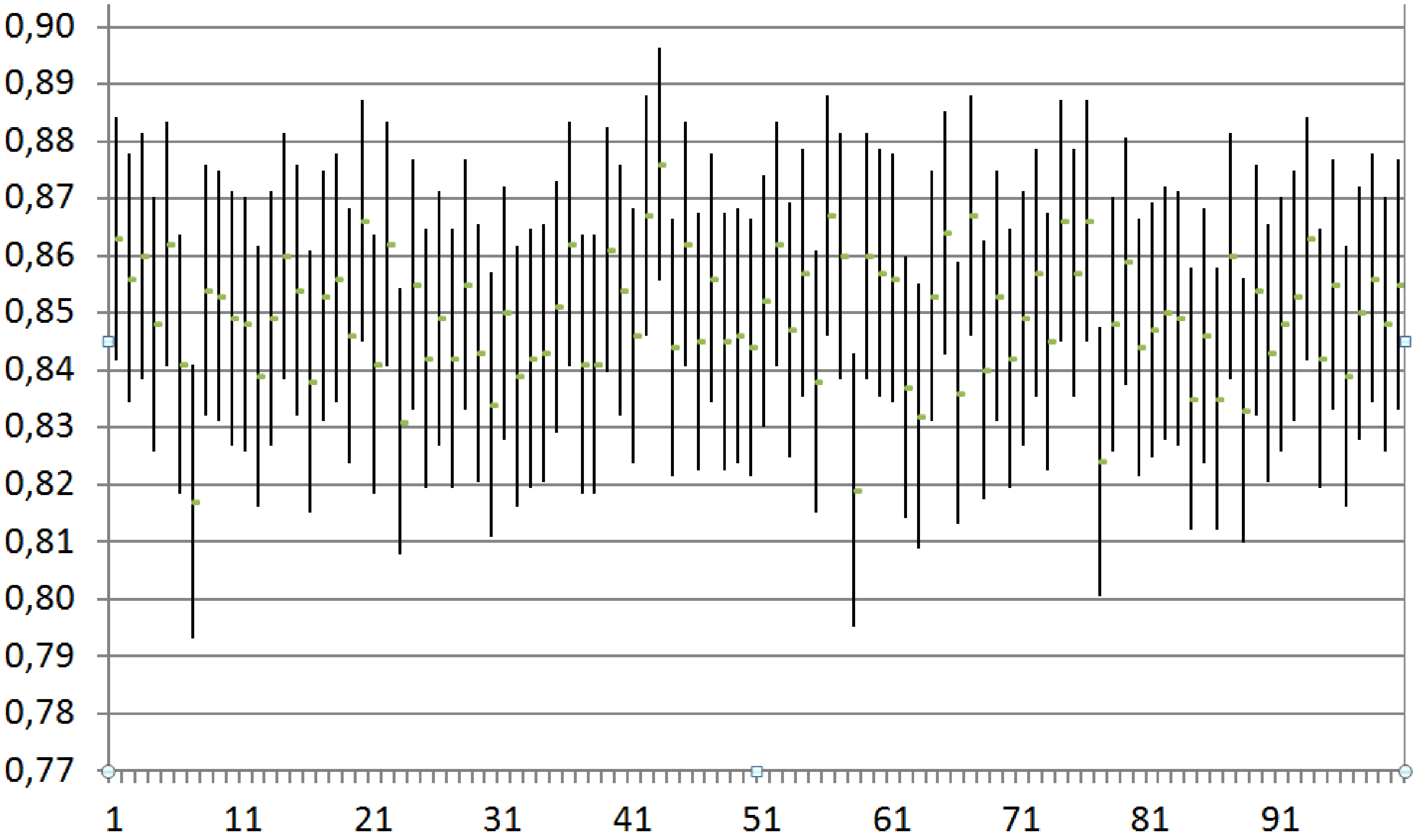
=NB.SI(B4:CW4;«OUI»)

=SI(ET(B1>=B5;B1<=B6); «OUI »;0)
(à dupliquer dans la plage B4:CW4).

=B7+1,96*RACINE(B7*(1-B7)/1000)
(à dupliquer dans la plage B6:CW6).

=NB.SI(B9:B1008;"=1")/1000
(à dupliquer dans la plage B7:CW7).

Représentation des 100 intervalles de confiance obtenus par simulation



Exploitation de la simulation

- Conjecturer la valeur de *p* rentrée en B1 (utilisation F9) et comparaison avec valeur rentrée.

- Donner du sens à la notion d'intervalle de confiance.

- Mettre en évidence qu'un intervalle de confiance peut ne pas contenir *p*.

- Mettre en évidence que deux intervalles de confiance peuvent être disjoints.

Conclusion:
deux objets différents pour des problématiques différentes

	Intervalle de fluctuation	Intervalle de confiance
Objectifs:	prendre une décision	• estimer une proportion inconnue • juger d'une évolution significative