PROBABILITÉS ET STATISTIQUES

Terminale

Spécialité

Région académique

Plan Académique de Formation 2019-2020

Un objectif général inchangé…

- **Modéliser** des situations avec des outils probabilistes
- Comprendre **les lois du hasard** : (loi des grands nombres)
- Maîtriser leurs **fluctuations.**
- Exploiter et interpréter des **simulations**
- <u>Pas de retour à la combinatoire et aux probabilités classiques.</u>

…mais une nouvelle optique

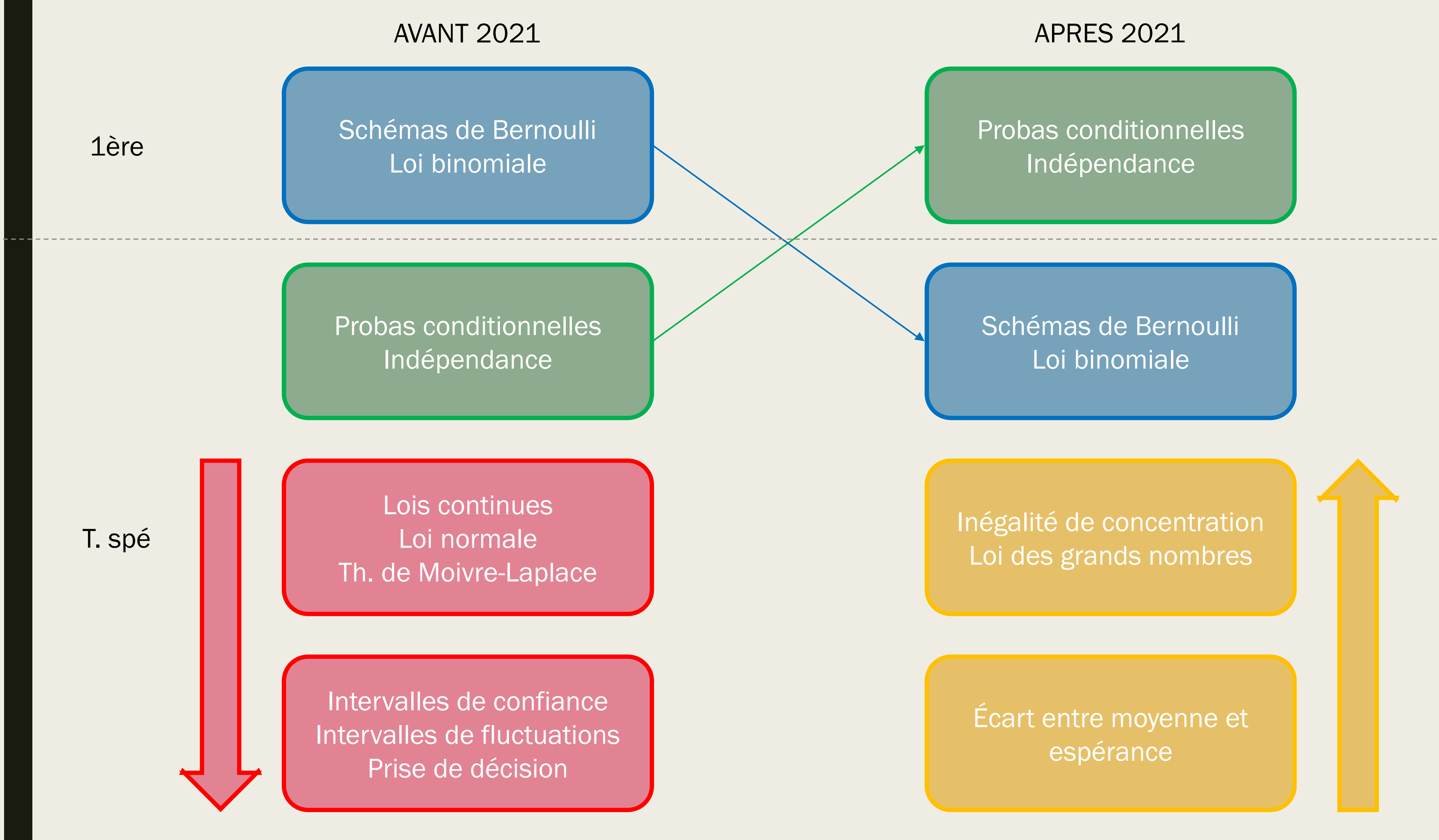
- **Le précédent programme de probabilité était centré sur les applications pratiques**
  - *médecine, sciences économiques et sociales, etc.*

- La spécialité s'adresse aux **futurs scientifiques**
  - *rigueur de la réflexion, abstraction, rôle de la démonstration, etc.*

→ **Élimination de ce qui n'est pas démontrable** et des simples applications
  - *variables à densité, Moivre-Laplace, prise de décision, etc.*

→ Focalisation sur les **démonstrations** pour un raisonnement complet

→ Appui sur la **simulation** pour donner du sens, mais pas pour justifier.


AVANT 2021
APRES 2021
1ère
T. spé
Schémas de Bernoulli
Loi binomiale
Probas conditionnelles
Indépendance
Probas conditionnelles
Indépendance
Schémas de Bernoulli
Loi binomiale
Lois continues
Loi normale
Th. de Moivre-Laplace
Inégalité de concentration
Loi des grands nombres
Intervalles de confiance
Intervalles de fluctuations
Prise de décision
Écart entre moyenne et espérance

Un exercice que l'on ne posera plus

Partie B

Cet ostréiculteur affirme que 60% de ses huîtres ont une masse supérieure à 91 g.
Un restaurateur souhaiterait lui acheter une grande quantité d'huîtres mais il voudrait, auparavant, vérifier l'affirmation de l'ostréiculteur.

Le restaurateur achète auprès de cet ostréiculteur 10 douzaines d'huîtres qu'on considèrera comme un échantillon de 120 huîtres tirées au hasard. Sa production est suffisamment importante pour qu'on l'assimile à un tirage avec remise.

Il constate que 65 de ces huîtres ont une masse supérieure à 91 g.

1. Soit F la variable aléatoire qui à tout échantillon de 120 huîtres associe la fréquence de celles qui ont une masse supérieure à 91 g.
Après en avoir vérifié les conditions d'application, donner un intervalle de fluctuation asymptotique au seuil de 95% de la variable aléatoire F.

2. Que peut penser le restaurateur de l'affirmation de l'ostréiculteur ?

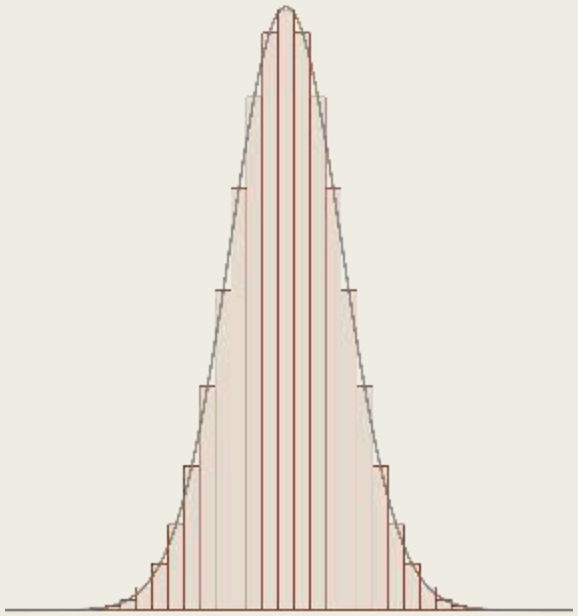
Source : Bac 2014 Antilles-Guyane, filière S.

Un théorème que l'on ne verra plus

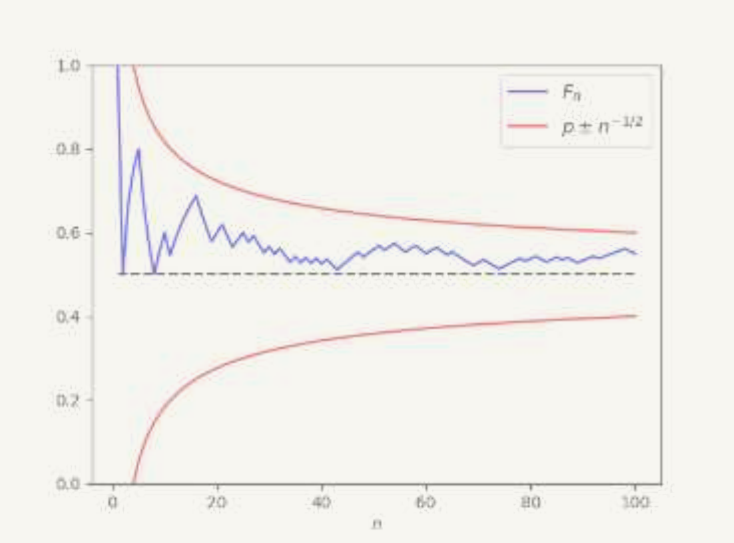
Le théorème de Moivre-Laplace

- Pourquoi ?
  - *Une « boîte noire » utilisée aveuglément,*
  - *La loi normale est inaccessible à la démonstration,*
  - *Le théorème de Moivre-Laplace encore plus,*
  - *Le sens échappait aux élèves.*

- Conséquences
  - *La loi normale (et les variables à densité) disparaissent des programmes*
  - *Les intervalles de confiance et de fluctuations* $\left[p-\frac{1}{\sqrt{n}}, p+\frac{1}{\sqrt{n}}\right], \left[p-u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p+u_\alpha\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right], \dots$ *disparaissent des programmes*



Structure des anciens Programmes



- Cycle 4 :
  - *On observait la* **stabilisation des fréquences** *(la fréquence F_n converge en probabilité vers p).*
- Seconde :
  - *On observait que* $P\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) > 0{,}95$ (aucun argument)
- 1ère
  - *On montrait que* nF_n *suit une loi binomiale* $B(n, p)$ (arbres)
- Terminale :
  - *On montrait que* F_n *suit asymptotiquement une loi normale* (Moivre-Laplace, admis)
  - *On en déduisait que* $P\left(F_n \in \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) \approx 1 - \alpha$

Structure des nouveaux programmes

- Cycle 4 :
  - *On observe la* ***stabilisation des fréquences***
    (la fréquence F_n converge en probabilité vers p).

- Seconde et 1ère : Mise en place du formalisme
  - *Arbres,*
  - *indépendance*
  - *Variables aléatoires finies*

- Terminale :
  - *On introduit la variance*
  - *On démontre l'inégalité de Tchebychev en général*
  - *On en déduit l'inégalité de concentration qui contrôle l'écart entre $\overline{X_n}$ et μ* (LGN)
  - *On applique cela au cas particulier de Bernoulli, où $\overline{X_n} = F_n$ et $\mu = p$.*

La variance

Un indicateur de dispersion relatif

Si X admet une espérance μ, la variance de X est

$$\boldsymbol{V(X) = E[(X-\mu)^2]}$$

- C'est un **indicateur de dispersion.**
- On peut **comparer** les fluctuations de deux variables aléatoires:
  - *Si V(X) > V(Y) alors X fluctue plus autour de sa moyenne que Y*
- Mais comment interpréter ses valeurs dans l'absolu ?
  - *Que peut-on dire si, par exemple,* $V(X) = 4$*?*

L’inégalité de Tchebychev
La variance comme indicateur absolu

- Inégalité de Tchebychev
  - *Si X admet une espérance μ et une variance σ^2, alors, pour tout $\epsilon > 0$,*

$$P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

- En particulier, pour tout entier k,

$$\boldsymbol{P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}}$$

- Exemple :
  - *Si $V(X) = 4$, alors on peut assurer que X s’écartera de sa moyenne de plus de $6 = 3\sigma$ avec une probabilité inférieure à $\frac{1}{9} \approx 11{,}1\%$.*

L'inégalité de concentration
Contrôler l'écart entre moyenne et espérance

- Soit X une variable aléatoire **d'espérance** μ et de variance σ^2

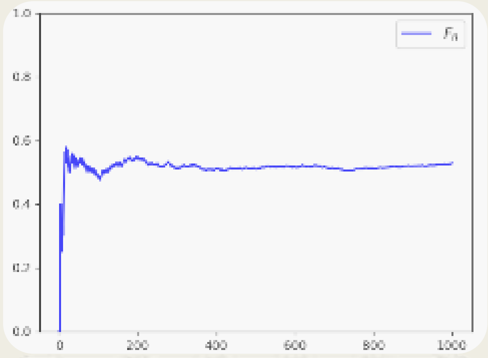
- Soit M_n la **moyenne** des valeurs sur un échantillon de taille n de X

$$E[M_n] = \mu \qquad V(M_n) = \frac{\sigma^2}{n}$$

- L'inégalité de Tchebychev devient **l'inégalité de concentration**

$$\boldsymbol{P(|M_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}}$$

On en déduit la LGN : M_n *converge en probabilité vers l'espérance* μ.



Le contrôle des fluctuations

Pour $\boldsymbol{k > 0}$, on a

$$\boldsymbol{P\left(|M_n - \mu| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}}$$

- **Exemple :** n lancers d'une pièce équilibré
  - *X de loi* $B\left(\frac{1}{2}\right) : \mu = p = \frac{1}{2}$ *et* $\sigma = \sqrt{\frac{1}{4}} = \frac{1}{2}$
  - $M_n = F_n$ *est la* ***fréquence*** *de Pile sur un échantillon de taille n de X*
  - *Avec k = 2 :*

$$\boldsymbol{P\left(\left|F_n - \frac{1}{2}\right| > \frac{1}{\sqrt{n}}\right) \leq \frac{1}{4}}$$

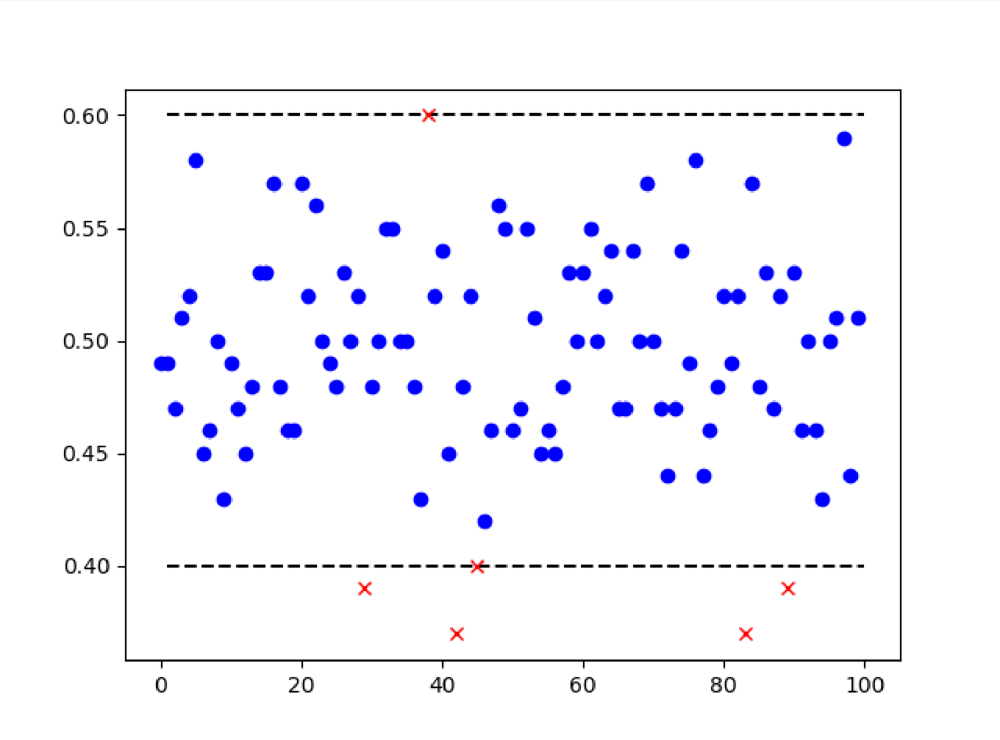
  - ***Conclusion :*** *Dans moins de 25% des cas, l'écart sur un échantillon de taille* n *entre la fréquence et la probabilité sera à plus de* $\frac{1}{\sqrt{n}}$.

Simulation

```python
from math import sqrt
from random import random

def ecartfreq(n=100,p=1/2,k=2):
    sigma = sqrt(p*(1-p))
    binf = p-k*sigma/sqrt(n)
    bsup = p+k*sigma/sqrt(n)
    N = 10000
    NbIn = 0
    for i in range(N):
        S = 0
        for j in range(n):
            if (random()<p):
                S = S+1
            f = S/n
        if (binf<f) and (f<bsup):
            NbIn = NbIn +1
    F = NbIn/N
    return F
```

```
In [31]: ecartfreq()
Out[31]: 0.9407
```



Les limites du nouveau programme
Moivre-Laplace en embuscade !

- **Observation :**
  - *95% des valeurs se trouvent à un écart inférieur à* $\frac{1}{\sqrt{n}}$
  - *L'inégalité de Tchebychev prévoyait au moins 75% des valeurs.*
  - *C'est correct mais imprécis*

- **Explication**
  - *L'inégalité de Tchebychev est **parfaitement générale**,*
  - *Elle est **loin d'être optimale** dans des cas précis,*
  - *Pour expliquer ces 95%, on est obligé de recourir à la **loi normale**,*
  - *Il en va de même dans la plupart des situations* (TCL)*.*

	Probabilités			Statistiques	
	Calcul	**Variables aléatoires**	**Lois**	**Descriptives**	**Inférentielles**
Cycle 4	Vocabulaire Équiprobabilité Tableaux croisés			Effectifs, fréquences Moyenne (non pondérée) Médiane Étendue	
Seconde	Formalisme ensembliste Crible Arbres de dénombrement			% de % Moyenne pondérée Écart interquartile Écart-type	Principe fréquentiste Fluctuations $p \pm \frac{1}{\sqrt{n}}$
Première	Conditionnement Indépendance Arbres pondérés Formule de Bayes	V.a. finies Loi Espérance Variance		Tableaux croisés	$\mathbf{P}(\lvert\overline{\mathbf{X}}_n - \mu\rvert \leq 2\sigma/\sqrt{n})$
T. spécialité	Schéma de Bernoulli	$\mathbf{E}[aX+Y]$ $\mathbf{var}(aX+Y)$ Inégalité de Tchebychev (*) $\mathbf{E}[XY]$	$\mathscr{B}(p)$ $\mathscr{B}(n,p)$ (*) $\mathscr{G}(p)$ (*) $\mathscr{P}(\lambda)$		Inégalité de concentration Loi des grands nombres Fluctuations pour $\mathscr{B}(n,p)$ $\mathbf{P}(\lvert\overline{\mathbf{X}}_n - \mu\rvert \leq k\sigma/\sqrt{n})$
T. comp.	Schéma de Bernoulli	V.a. à densité F.d.r. Espérance Variance	$\mathscr{U}([\![1,n]\!])$ $\mathscr{B}(p)$ $\mathscr{B}(n,p)$ $\mathscr{G}(p)$ $\mathscr{U}([a,b])$ $\mathscr{E}(\lambda)$	Nuages de points Moindres carrés	Fluctuations pour $\mathscr{B}(n,p)$
T. exp.	Marches aléatoires	Chaînes de Markov			

Tableau C.1 – Organisation des principales notions apparaissant dans les programmes 2019. Les notions marquées d'une astérisque sont celles qui sont proposées en approfondissement.