Association des Professeurs de Mathématiques de l'Enseignement Public
de Haute Normandie

La régionale APMEP de Haute-Normandie

vous invite à participer à la journée régionale

Mercredi 2 octobre 2013

Au lycée Marcel Sembat, Sotteville-lès-Rouen

Atelier : Probabilités et statistiques en terminale

avec Sylvie Colesse et Christian Vassard

Les probabilités et les statistiques ont fait une entrée en force dans les nouveaux programmes de lycée. Après une première année d'enseignement en terminale, l'heure est maintenant à l'analyse et au recul. À partir d'exemples concrets et en nous appuyant sur l'outil informatique, nous nous proposons dans cet atelier d'apporter un éclairage sur les différentes notions introduites et une réflexion sur les concepts qui les sous-tendent. Nous mettrons notamment en évidence les liens qui unissent la loi binomiale à la loi normale, les intervalles de fluctuation aux intervalles de confiance. Nous demandons à chaque participant de venir avec une calculatrice graphique type lycée.

PONDICHERY 2013 (S)

Dans une entreprise, on s'intéresse à la probabilité qu'un salarié soit absent durant une période d'épidémie de grippe.

1. (...)

2. (...)

3. Cette entreprise emploie 220 salariés. Pour la suite on admet que la probabilité pour qu'un salarié soit malade une semaine donnée durant cette période d'épidémie est égale à $p = 0,05$.
On suppose que l'état de santé d'un salarié ne dépend pas de l'état de santé de ses collègues.
On désigne par X la variable aléatoire qui donne le nombre de salariés malades une semaine donnée.

a. Justifier que la variable aléatoire X suit une loi binomiale dont on donnera les paramètres.
Calculer l'espérance mathématique μ et l'écart type σ de la variable aléatoire X.

b. On admet que l'on peut approcher la loi de la variable aléatoire $\dfrac{X-\mu}{\sigma}$ par la loi normale centrée réduite c'est-à-dire de paramètres 0 et 1.
On note Z une variable aléatoire suivant la loi normale centrée réduite.
Le tableau suivant donne les probabilités de l'évènement $Z < x$ pour quelques valeurs du nombre réel x :

x	$-1,55$	$-1,24$	$-0,93$	$-0,62$	$-0,31$	$0,00$	$0,31$	$0,62$	$0,93$	$1,24$	$1,55$
$P(Z < x)$	$0,061$	$0,108$	$0,177$	$0,268$	$0,379$	$0,500$	$0,621$	$0,732$	$0,823$	$0,892$	$0,939$

Calculer, au moyen de l'approximation proposée en question b, une valeur approchée à 10^{-2} près de la probabilité de l'événement : «le nombre de salariés absents dans l'entreprise au cours d'une semaine donnée est supérieur ou égal à 7 et inférieur ou égal à 15».

Extrait du document d’accompagnement de terminale S sur probabilité et statistiques

Remarque

Quand $n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$, il est courant de faire les calculs impliquant une variable binomiale en la remplaçant par une variable suivant une loi normale de mêmes espérance et variance.

Seul le programme de STI2D-STL mentionne cette pratique, qui ne doit donc pas être mise en œuvre dans les autres filières où tous les calculs de probabilités se font à la calculatrice en utilisant la loi exacte (au programme), quelle qu'elle soit.

Les calculs d'intervalles de fluctuation et d'intervalles de confiance se font avec les formules données dans le programme.

Ministère de l'éducation nationale, de la jeunesse et de la vie associative (DGESCO) **Page 21 sur 70**

Mathématiques - Probabilités et statistique

http://eduscol.education.fr/prog

Loi normale et loi binomiale

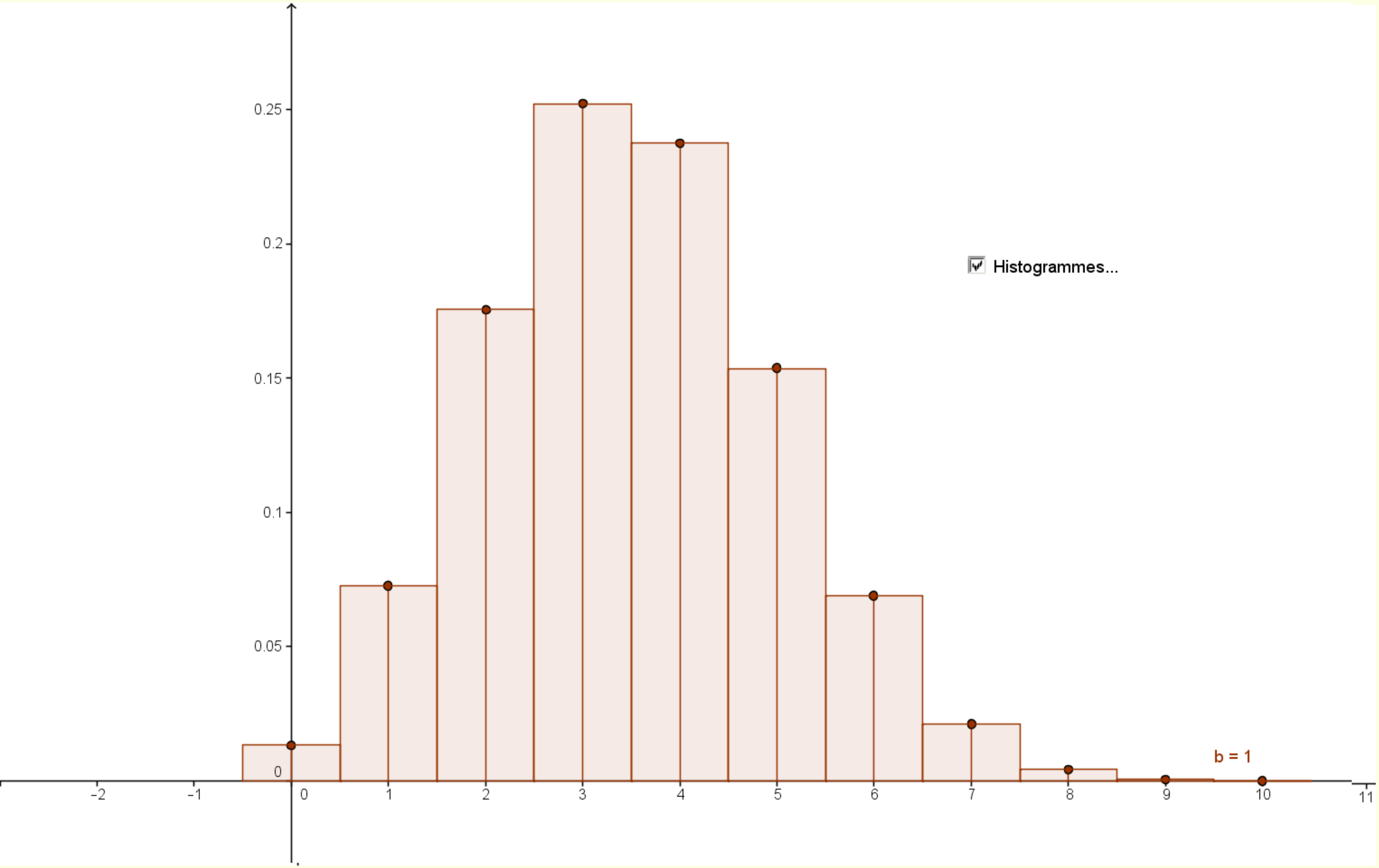
La loi binomiale est une loi conceptuellement simple mais qui conduit vite à des calculs inextricables, notamment à la main mais aussi avec un outil de calcul moderne

On jette un dé non pipé 12000 fois. Quelle est la probabilité d'avoir entre 1950 et 2050 fois le numéro 6?

On jette un dé 1 200 000 fois. Quelle est la probabilité d'avoir entre 199 000 et 201 000 fois le numéro 6?

Quelle loi continue peut approximer la loi binomiale?

Des diagrammes en bâtons vers un histogramme… (fichier)



Qu’observe-t-on? (fichier)

Centrer et réduire la loi binomiale…

Pour un réel p dans l'intervalle]0,1[et un entier naturel n, on note X_n la variable aléatoire qui suit la loi binomiale de paramètres n et p.

X_n prend ses valeurs dans $\{0, 1, \ldots, n\}$ et $\boxed{E(X_n) = np\ ,\ V(X_n) = np(1-p)\ ,\ s(X) = \sqrt{np(1-p)} = s}$

La variable centrée réduite associée à X_n est donc $Z_n = \dfrac{X_n - np}{s}$

C'est une variable discrète d'espérance 0 et d'écart-type 1.

La variable Z_n prend les valeurs $\dfrac{k - np}{s}$ pour k entier allant de 0 à n.

Ces valeurs ne sont pas entières mais réparties régulièrement entre $\dfrac{-np}{s}$ et $\dfrac{n(1-p)}{s}$

l'écart entre deux valeurs consécutives étant $\dfrac{1}{s}$.

Fichier ggb

Tiré du document d'accompagnement :

Théorème (de Moivre-Laplace)

On suppose que, pour tout entier n, la variable aléatoire X_n suit une loi binomiale $\mathcal{B}(n, p)$.

On pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$, variable centrée et réduite associée à X_n .

Alors, pour tous réels a et b tels que $a < b$, on a : $\lim\limits_{n \to +\infty} \mathrm{P}(a \leq Z_n \leq b) = \int_a^b \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$.

Voici ce que dit Laplace à propos des travaux de Moivre :

« Moivre a repris dans son ouvrage [*The doctrine of Chances*] le théorème de Jacques Bernoulli sur la probabilité des résultats déterminés par un grand nombre d'observations[2].

Il ne se contente pas de faire voir, comme Bernoulli, que le rapport des événements qui doivent arriver approche sans cesse de celui de leurs possibilités respectives, il donne de plus une expression élégante et simple de la probabilité que la différence de ces deux rapports soit contenue dans des limites données. »

Plus loin avec l’approximation proposée dans l’exercice…

Correction de continuité

Ce qui revient d’ailleurs à approximer $P(X=7)$ par $\int_{6,5}^{7,5} f(x)dx$

et non par $\int_{7}^{7} f(x)dx$ … qui vaut 0…

Soit *X* une variable aléatoire réelle d'espérance mathématique μ et d'écart-type σ.
La variable aléatoire centrée réduite *R* associée à *X* est définie par:

$$R = \frac{X - m}{s}$$

Il est bien clair que :

$$E(R) = 0 \qquad \text{et} \qquad s(R) = 1$$

Centrée *Réduite*

Centrer , réduire: un exemple…

On peut aussi envisager l’exemple des notes à un contrôle dans une classe.

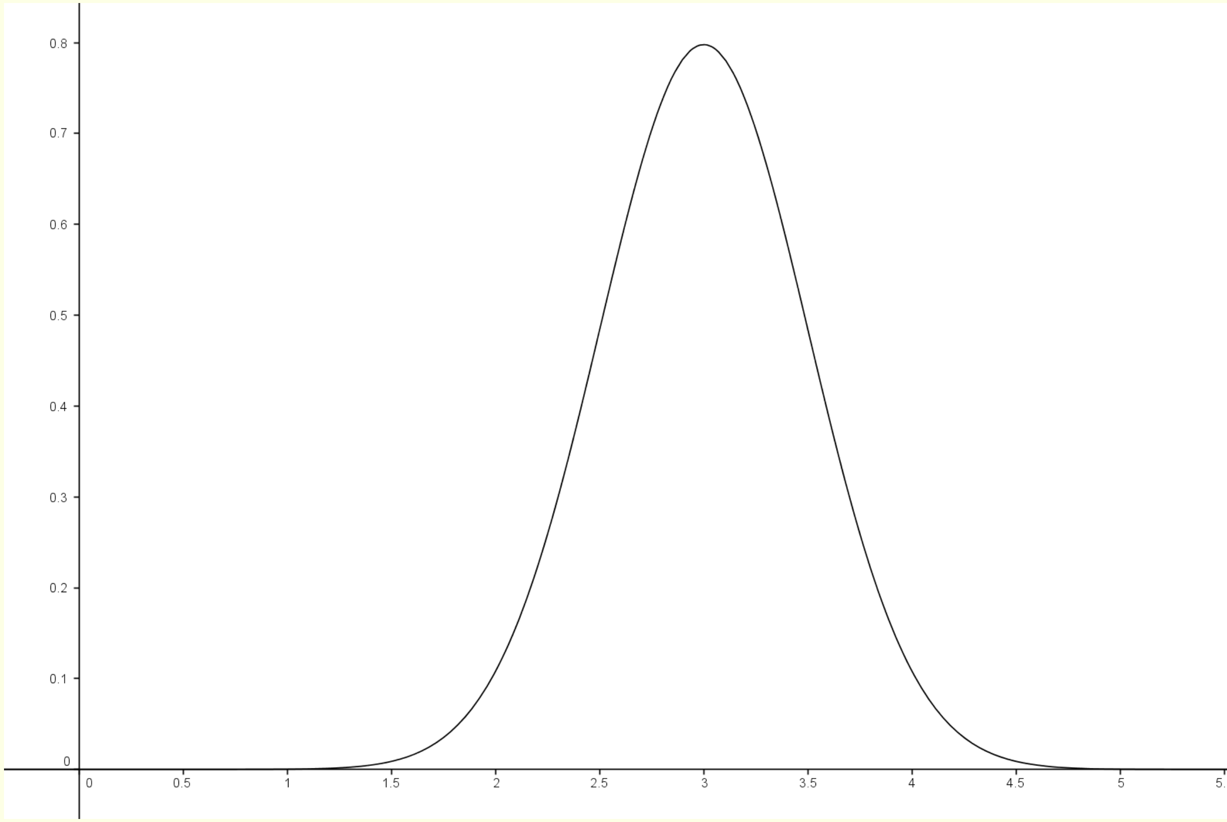
Quelle est la meilleure « performance »?

Un élève qui a 10,2 avec un moyenne et un écart-type de classe de 9 et 3,5…

Ou un élève qui a 9,3 avec un moyenne et un écart-type de classe de 8 et 1,9?

Définition

Une variable aléatoire X suit une loi normale de paramètres μ et σ^2, notée $N(\mu, \sigma^2)$ lorsque la variable aléatoire centrée réduite associée $Y = (X - \mu)/\sigma$ suit la loi normale $N(0,1)$.



On a :

$$X = \sigma Y + \mu$$

D'où l'on déduit que

$E(X) = \mu$ et $V(X) = \sigma^2$.

Retour sur le théorème de Moivre Laplace

X_n désigne une loi binomiale d'espérance μ_n et d'écart-type σ_n ; N est la loi normale centrée réduite. Alors

$$P(a \le X_n \le b) = P\left(\frac{a-\mu_n}{\sigma_n} \le \frac{X-\mu_n}{\sigma_n} \le \frac{b-\mu_n}{\sigma_n}\right)$$

$$\approx P\left(\frac{a-\mu_n}{\sigma_n} \le N \le \frac{b-\mu_n}{\sigma_n}\right) = P(a \le Y_n = \sigma_n N + \mu_n \le b)$$

Il est alors clair que la variable aléatoire Y_n suit une loi normale de paramètre μ_n et σ_n^2

On peut donc directement approximer la loi binomiale par la loi normale qui possède la même espérance et la même variance.

LIBAN 2013 (S)

L'entreprise *Fructidoux* fabrique des compotes qu'elle conditionne en petits pots de 50 grammes. Elle souhaite leur attribuer la dénomination «compote allégée».

La législation impose alors que la teneur en sucre, c'est-à-dire la proportion de sucre dans la compote, soit comprise entre 0,16 et 0,18. On dit dans ce cas que le petit pot de compote est conforme.

L'entreprise possède deux chaînes de fabrication F_1 et F_2.

Les parties A, B et C peuvent être traitées indépendamment les unes des autres.

Partie A (...)

Partie B

1. On note X la variable aléatoire qui, à un petit pot pris au hasard dans la production de la chaîne F_1 associe sa teneur en sucre.

   On suppose que X suit la loi normale d'espérance $m_1 = 0,17$ et d'écart-type $\sigma_1 = 0,006$. Dans la suite, on pourra utiliser le tableau ci-dessous :

α	β	$P(\alpha \leq X \leq \beta)$
$0,13$	$0,15$	$0,0004$
$0,14$	$0,16$	$0,0478$
$0,15$	$0,17$	$0,4996$
$0,16$	$0,18$	$0,9044$
$0,17$	$0,19$	$0,4996$
$0,18$	$0,20$	$0,0478$
$0,19$	$0,21$	$0,0004$

Donner une valeur approchée à 10^{-4} près de la probabilité qu'un petit pot prélevé au hasard dans la production de la chaîne F_1 soit conforme.

2. On note Y la variable aléatoire qui, à un petit pot pris au hasard dans la production de la chaîne F_2, associe sa teneur en sucre.

On suppose que Y suit la loi normale d'espérance $m_2 = 0,17$ et d'écart-type σ_2.

On suppose de plus que la probabilité qu'un petit pot prélevé au hasard dans la production de la chaîne F_2 soit conforme est égale à $0,99$.

Soit Z la variable aléatoire définie par $Z = \dfrac{Y - m_2}{\sigma_2}$.

a. Quelle loi la variable aléatoire Z suit-elle ?

b. Déterminer, en fonction de σ_2 l'intervalle auquel appartient Z lorsque Y appartient à l'intervalle [0,16 ; 0,18].

c. En déduire une valeur approchée à 10^{-3} près de σ_2.

On pourra utiliser le tableau donné ci-dessous, dans lequel la variable aléatoire Z suit la loi normale d'espérance 0 et d'écart-type 1 :

β	$P(-\beta \leq Z \leq \beta)$
$2,4324$	$0,985$
$2,4573$	$0,986$
$2,4838$	$0,987$
$2,5121$	$0,988$
$2,5427$	$0,989$
$2,5758$	$0,990$
$2,6121$	$0,991$
$2,6521$	$0,992$
$2,6968$	$0,993$

CENTRES ETRANGERS 2013 (S)

Partie A et B

(...)

Partie C

L'industriel affirme que seulement 2 % des vannes qu'il fabrique sont défectueuses. On suppose que cette affirmation est vraie, et l'on note F la variable aléatoire égale à la fréquence de vannes défectueuses dans un échantillon aléatoire de 400 vannes prises dans la production totale.

1. Déterminer l'intervalle I de fluctuation asymptotique au seuil de 95 % de la variable F,
2. On choisit 400 vannes au hasard dans la production, On assimile ce choix à un tirage aléatoire de 400 vannes, avec remise, dans la production.

   Parmi ces 400 vannes, 10 sont défectueuses.

   Au vu de ce résultat peut-on remettre en cause. au seuil de 95 %, l'affirmation de l'industriel ?

Partie D

(...)

Intervalles de fluctuation et prise de décision

Théorème

Si la variable aléatoire X_n suit la loi binomiale de paramètres n et p, avec p dans l'intervalle]0 ; 1[, alors pour tout réel α dans l'intervalle]0 ; 1[, on a:

$$\lim_{n \to +\infty} P\left(\frac{X_n}{n} \in I_n\right) = 0,95$$

où I_n désigne l'intervalle $\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$

On pose parfois $F_n = \frac{X_n}{n}$; F_n correspond à la fréquence du succès pour la loi binomiale considérée.

Démonstration (ici exigible en terminale S)

D'après le théorème de Moivre-Laplace, on a $\lim\limits_{n \to +\infty} P(-1{,}96 \leq Z_n \leq 1{,}96) = P(-1{,}96 \leq Z \leq 1{,}96)$

$$= 0{,}95$$

où $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$ et où Z suit la loi normale centrée réduite.

Regardons maintenant le premier membre et plus précisément l'expression dont on prend la limite:

$$P(-1{,}96 \leq Z_n \leq 1{,}96) = P\left(-1{,}96 \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq 1{,}96\right)$$

$$= P\left(np - 1{,}96\sqrt{np(1-p)} \leq X_n \leq np + 1{,}96\sqrt{np(1-p)}\right)$$

$$= P\left(p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$$

Autrement dit :

$$\boxed{\lim_{n \to +\infty} P\left(p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 0{,}95.}$$

On admet donc que, sous certaines conditions, notamment pour n suffisamment grand on peut approcher le terme de rang *n* de la suite

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

par sa limite 0,95.

Visualisation à l'aide de geogebra

Les conditions communément admises pour pratiquer cette approximation sont:

$n \geq 30$, $np \geq 5$ et $n(1 - p) \geq 5$.

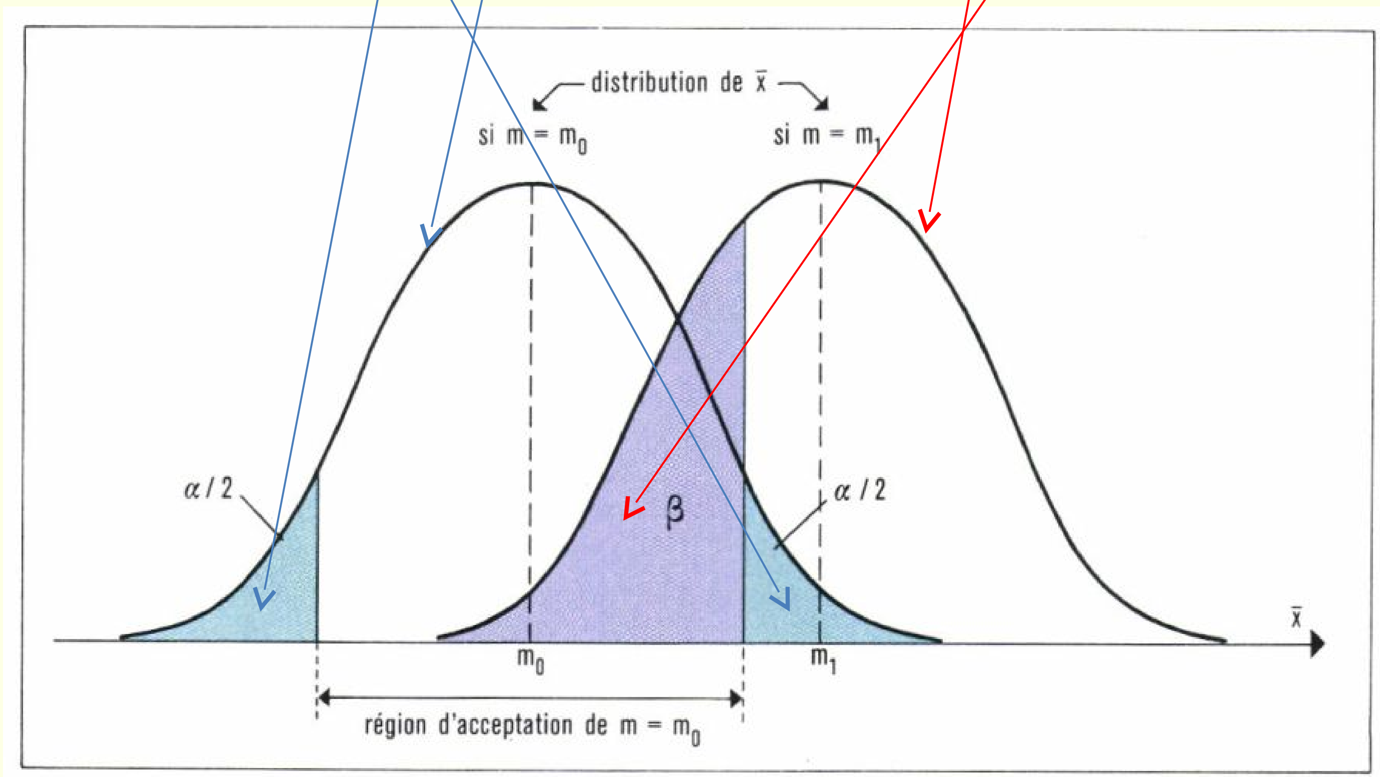
Au seuil de 95%, on prendra

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

Au seuil de 99%, on prendra

$$\left[p - 2,58\frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 2,58\frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

décision \ réalité	H_0 vraie	H_0 fausse
non-rejet de H_0	correct	manque de puissance (risque de deuxième espèce)
rejet de H_0	rejet à tort (risque de première espèce)	correct



Test bilatéral de l'hypothèse $m = m_0$, au risque α. $1 - \beta$ = puissance du test si $m = m_1$.

ANTILLES GUYANE 2013 (S)

Partie A

Soient n un entier naturel, p un nombre réel compris entre 0 et 1 et X_n une variable aléatoire suivant une loi binomiale de paramètres n et p. On note $F_n = \frac{X_n}{n}$ et f une valeur prise par F_n. On rappelle que, pour n assez grand, l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ contient la fréquence f avec une probabilité au moins égale à $0,95$.

En déduire que l'intervalle $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ contient p avec une probabilité au moins égale à $0,95$.

Partie B

On cherche à étudier le nombre d'étudiants connaissant la signification du sigle URSSAF. Pour cela, on les interroge en proposant un questionnaire à choix multiples. Chaque étudiant doit choisir parmi trois réponses possibles, notées A, B et C, la bonne réponse étant la A.

On note r la probabilité pour qu'un étudiant connaisse la bonne réponse. Tout étudiant connaissant la bonne réponse répond A, sinon il répond au hasard (de façon équiprobable).

1. On interroge un étudiant au hasard. On note :

A l'évènement «l'étudiant répond A»,

B l'évènement «l'étudiant répond B»,

C l'évènement «l'étudiant répond C»,

R l'évènement «l'étudiant connait la réponse»,

$\overline{R}$ l'évènement contraire de R.

a. Traduire cette situation à l'aide d'un arbre de probabilité.

b. Montrer que la probabilité de l'évènement A est $P(A) = \frac{1}{3}(1 + 2r)$.

c. Exprimer en fonction de r la probabilité qu'une personne ayant choisie A connaisse la bonne réponse.

2. Pour estimer r, on interroge 400 personnes et on note X la variable aléatoire comptant le nombre de bonnes réponses. On admettra qu'interroger au hasard 400 étudiants revient à effectuer un tirage avec remise de 400 étudiants dans l'ensemble de tous les étudiants.

a. Donner la loi de X et ses paramètres n et p en fonction de r. On utilisera le fait que $p = P(A) = \frac{1}{3}(1 + 2r)$ démontré dans la question précédente.

b. Dans un premier sondage, on constate que 240 étudiants répondent A, parmi les 400 interrogés.
Donner un intervalle de confiance au seuil de 95 % de l'estimation de p.
En déduire un intervalle de confiance au seuil de 95 % de r.

c. Dans la suite, on suppose que $r = 0,4$. Compte-tenu du grand nombre d'étudiants, on considérera que X suit une loi normale.

i. Donner les paramètres de cette loi normale.

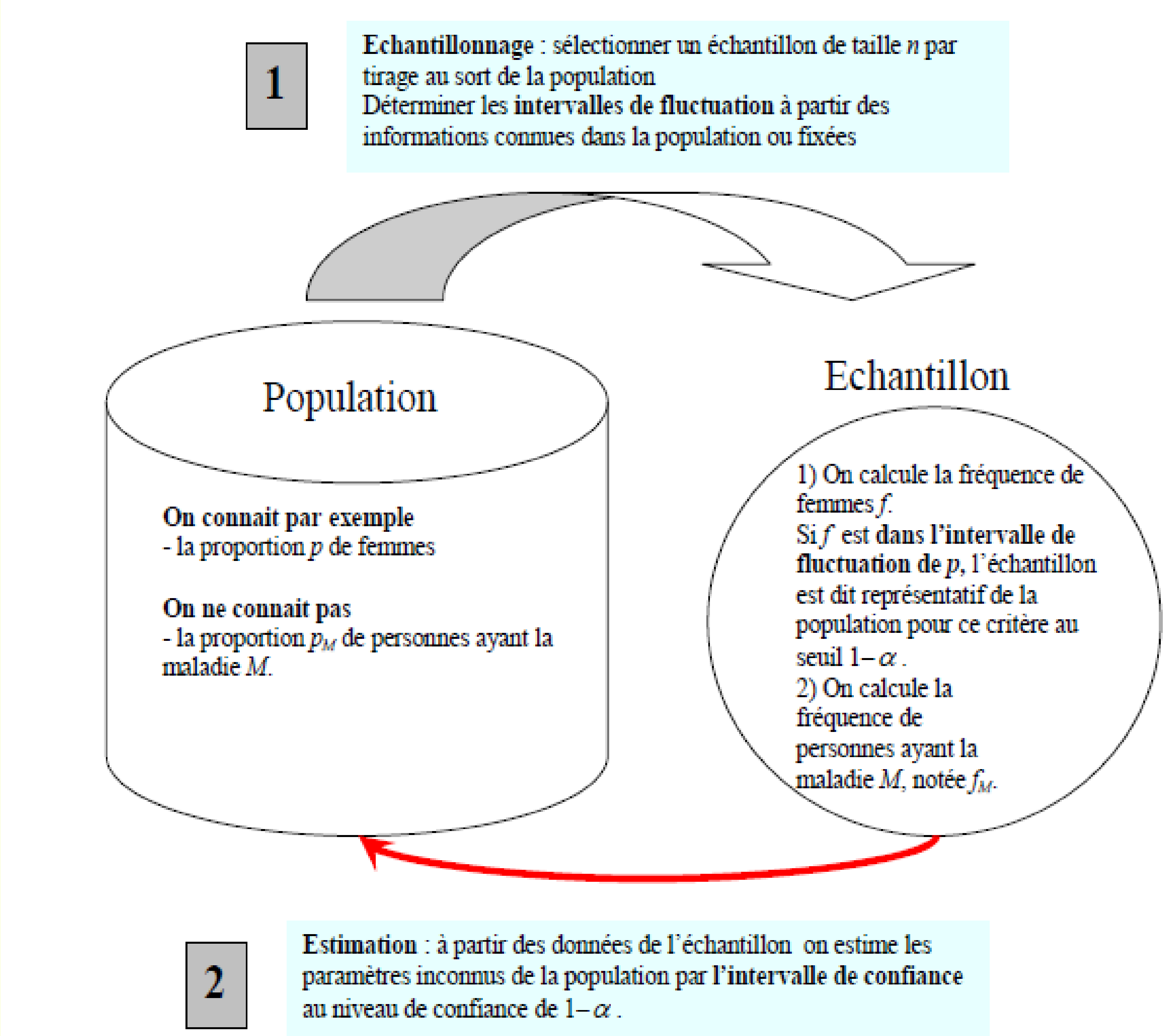
ii. Donner une valeur approchée de $P(X \leqslant 250)$ à 10^{-2} près.
On pourra s'aider de la table suivante, qui donne une valeur approchée de $P(X \leqslant t)$ où X est la variable aléatoire de la question **2. c**.

E12					=LOI.NORMALE($A12+E$1 ;240 ;RACINE(96) ;VRAI)						
	A	B	C	D	E	F	G	H	I	J	K
1	t	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
2	235	0,305	0,309	0,312	0,316	0,319	0,323	0,327	0,330	0,334	0,338
3	236	0,342	0,345	0,349	0,353	0,357	0,360	0,364	0,368	0,372	0,376
4	237	0,380	0,384	0,388	0,391	0,395	0,399	0,403	0,407	0,411	0,415
5	238	0,419	0,423	0,427	0,431	0,435	0,439	0,443	0,447	0,451	0,455
6	239	0,459	0,463	0,467	0,472	0,476	0,480	0,484	0,488	0,492	0,496
7	240	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,533	0,537
8	241	0,541	0,545	0,549	0,553	0,557	0,561	0,565	0,569	0,573	0,577
9	242	0,581	0,585	0,589	0,593	0,597	0,601	0,605	0,609	0,612	0,616
10	243	0,620	0,624	0,628	0,632	0,636	0,640	0,643	0,647	0,651	0,655
11	244	0,658	0,662	0,666	0,670	0,673	0,677	0,681	0,684	0,688	0,691
12	245	0,695	0,699	0,702	**0,706**	0,709	0,713	0,716	0,720	0,723	0,726
13	246	0,730	0,733	0,737	0,740	0,743	0,746	0,750	0,753	0,756	0,759
14	247	0,763	0,766	0,769	0,772	0,775	0,778	0,781	0,784	0,787	0,790
15	248	0,793	0,796	0,799	0,802	0,804	0,807	0,810	0,813	0,815	0,818
16	249	0,821	0,823	0,826	0,829	0,831	0,834	0,836	0,839	0,841	0,844
17	250	0,846	0,849	0,851	0,853	0,856	0,858	0,860	0,863	0,865	0,867
18	251	0,869	0,871	0,874	0,876	0,878	0,880	0,882	0,884	0,886	0,888
19	252	0,890	0,892	0,893	0,895	0,897	0,899	0,901	0,903	0,904	0,906
20	253	0,908	0,909	0,911	0,913	0,914	0,916	0,917	0,919	0,921	0,922
21	254	0,923	0,925	0,926	0,928	0,929	0,931	0,932	0,933	0,935	0,936
22	255	0,937	0,938	0,940	0,941	0,942	0,943	0,944	0,945	0,947	0,948
23	256	0,949	0,950	0,951	0,952	0,953	0,954	0,955	0,956	0,957	0,958
24	257	0,959	0,960	0,960	0,961	0,962	0,963	0,964	0,965	0,965	0,966
25	258	0,967	0,968	0,968	0,969	0,970	0,970	0,971	0,972	0,972	0,973
26	259	0,974	0,974	0,975	0,976	0,976	0,977	0,977	0,978	0,978	0,979
27	260	0,979	0,980	0,980	0,981	0,981	0,982	0,982	0,983	0,983	0,984

Extrait d'une feuille de calcul

-

Exemple d'utilisation :
au croisement de la ligne 12 et de la colonne E le nombre $0,706$ correspond à $P(X \leqslant 245,3)$.


1
Echantillonnage : sélectionner un échantillon de taille n par tirage au sort de la population
Déterminer les **intervalles de fluctuation** à partir des informations connues dans la population ou fixées
Population
On connait par exemple
- la proportion p de femmes
On ne connait pas
- la proportion p_M de personnes ayant la maladie M.
Echantillon
1) On calcule la fréquence de femmes f.
Si f **est dans l'intervalle de fluctuation de** p, l'échantillon est dit représentatif de la population pour ce critère au seuil $1-\alpha$.
2) On calcule la fréquence de personnes ayant la maladie M, notée f_M.
2
Estimation : à partir des données de l'échantillon on estime les paramètres inconnus de la population par **l'intervalle de confiance** au niveau de confiance de $1-\alpha$.

On redonne en terminale l'intervalle de confiance de seconde:

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

On peut aussi connaître cet intervalle, valable pour n grand et f pas trop petit:

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

Intervalle de fluctuation ?

La proportion *p* dans la population est connue ou encore on fait une hypothèse sur sa valeur.

Intervalle de confiance ?

On cherche à estimer une proportion *p* inconnue

Fichier geogebra

Probabilité ou confiance ?

Avant le tirage d’un échantillon, la procédure d’obtention de l’intervalle de confiance a une probabilité 1 – α que cet intervalle contienne le paramètre inconnu.

Après le tirage, le paramètre est dans l’intervalle calculé avec une confiance de 1 – α.

L’expression « *p* a 95% de chances d’appartenir à un intervalle de confiance comme [0,504 ; 0,696] » n’a pas de sens: ***cette expression ne contient plus rien d’aléatoire***, *p* est ou non dans cet intervalle sans que le hasard n’intervienne.
Mais on sait que si on répète l’estimation un grand nombre de fois, 95% des intervalles de confiance générés contiendront la probabilité cherchée.
C’est là où se niche la confiance.

$$P\left(F-\frac{1}{\sqrt{n}}\leq p\leq F+\frac{1}{\sqrt{n}}\right)\geq P\left(F-1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\leq p\leq F+1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)=0{,}95$$

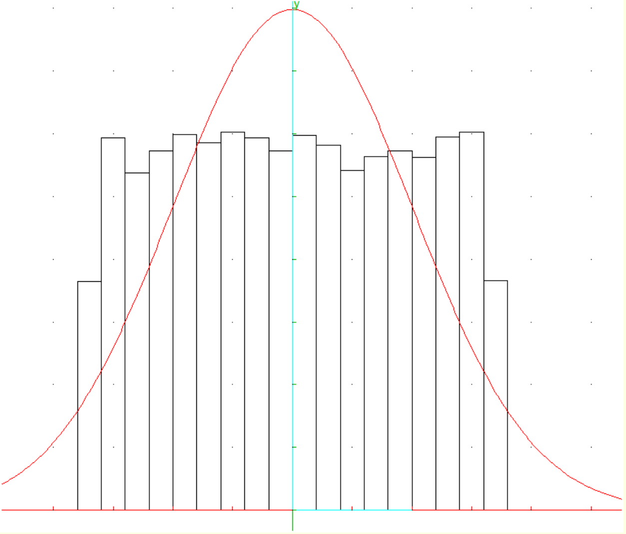
On a alors une probabilité supérieure ou égale à 0,95 d'exhiber une fourchette contenant la valeur p à estimer.

Le théorème central limite

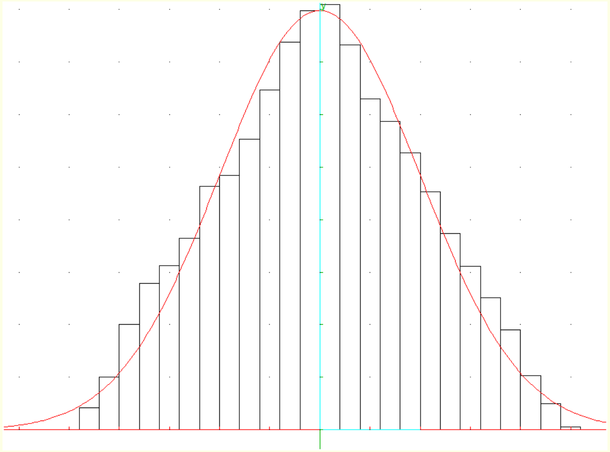
Soit X_1, X_2,…, X_n une suite de variables aléatoires indépendantes et identiquement distribuées (c'est-à-dire ayant toutes la même loi de probabilité), possédant une moyenne μ et une variance σ^2 finie, alors quand n tend vers l'infini, la variable aléatoire

$X_T = X_1 + \dots + X_n$ est telle que $\dfrac{X_T - n\mu}{s\sqrt{n}}$ converge en probabilité vers la loi normale

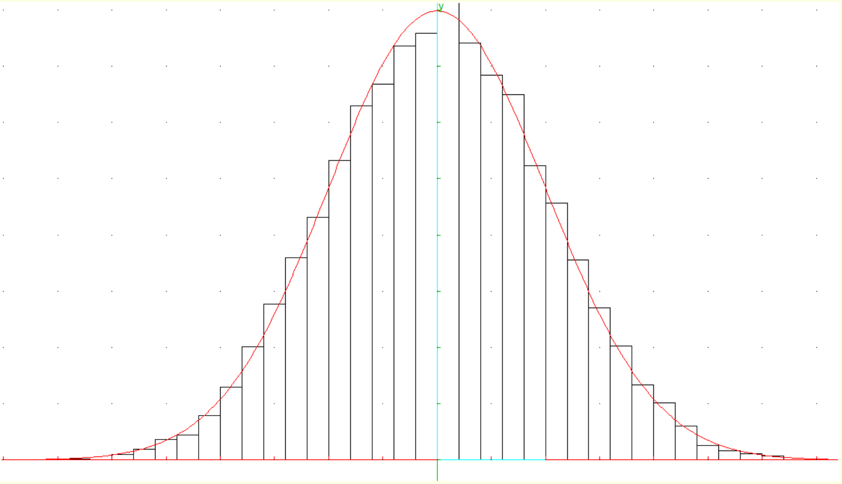
centrée réduite.



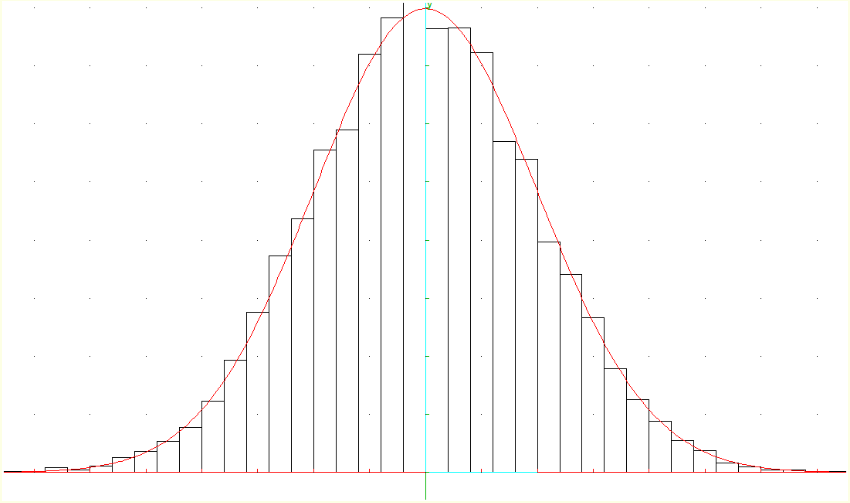
n=1



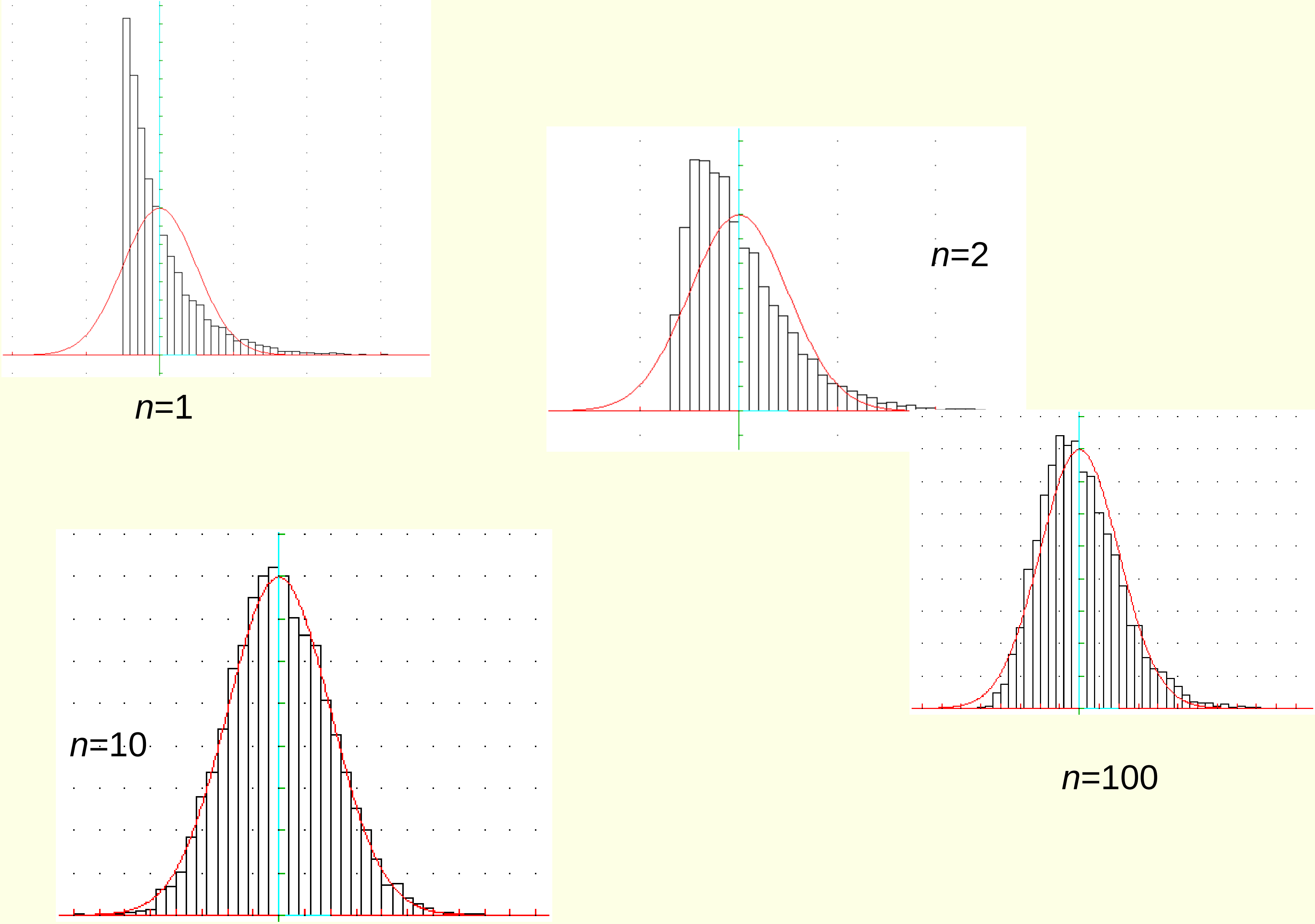
n=2



n=10



n=100


n=1
n=2
n=10
n=100