STAGE
STATISTIQUE ET PROBABILITES

30-31 AOUT 2012

Frédéric Lavancier
David Simon

PLAN

Comparatif des fonctionnalités des logiciels

		Open Office	Xcas	GeoGebra	Algobox	Scilab	R	Calculatrices Usuelles
Calcul de proba et proba cumulé :								
	Binomiale	OK	N<35	OK	N<70	OK	OK	OK
	Uniforme	Non	Non	OK	Non	Non	OK	OK
	Exponentielle	OK	Non	OK	Non	OK	OK	Non
	Normale	OK	OK	OK	OK	OK	OK	OK
	Hypergéométrique	OK	Non	OK	Non	Non	OK	OK
Calcul de quantile (proba inverse)								
	Binomiale	OK	N<35	OK	Non	OK	OK	OK
	Uniforme	Non	Non	Non	Non	Non	OK	Non
	Exponentielle	Non	Non	OK	Non	OK	OK	Non
	Normale	OK	OK	OK	OK	OK	OK	OK
	Hypergéométrique	Non	Non	OK	Non	Non	OK	OK
Simulation de valeurs selon la loi								
	Binomiale	Non	Non	OK	Non	OK	OK	OK
	Uniforme	OK	OK	OK	OK	OK	OK	OK
	Exponentielle	Non	Non	Non	Non	OK	OK	Non
	Normale	Non	OK	OK	Non	OK	OK	OK
	Hypergéométrique	Non	Non	Non	Non	Non	OK	Non
Tirer un échantillon de valeurs dans un vecteur (sondage)		Non	OK	OK	Non	OK	OK	Non
Importer des données		OK	Non	OK	Non	OK	OK	Non
Histogramme		pas direct	pas direct	OK	Non	OK	OK	OK
Diagramme en batons		OK	Non	pas direct	Non	OK	OK	Non
Diagramme en boîte (boîte à moustaches)		Non	OK	OK	Non	OK	OK	OK
Calcul formel		Non	OK	OK	Non	Non	Non	Non
Approximation numérique (d'intégrales)		Non	OK	OK	Non	OK	OK	OK
Programmation		Basique	OK	Basique	OK	OK	OK	OK
Environnement “clique-boutons”		OK	Non	OK	Non	Non	OK (R Commander)	Non

Partie I : Loi Binomiale.

I. 1. Calculs de probabilités

Exercice 1 :. Application directe (p.250-251 repères, 1ère S, Hachette)

76. Soit X une variable aléatoire qui suit une loi binomiale de paramètres 6 et $\frac{3}{4}$.

1. Donner, sous forme de fraction irréductible, l'expression de $p(X=0)$, $p(X=1)$ et $p(X=2)$.
2. En déduire $p(X \geqslant 3)$.

81. Un fabricant vend des stylos par lot de 10. Soit X la variable aléatoire correspondant au nombre de stylos défectueux dans un lot. On admet que X suit une loi binomiale de paramètres 10 et 0,034.

On donnera les résultats à 10^{-3} près.

1. Quelle est la probabilité que le lot contienne au moins un stylo défectueux ?
2. Quelle est la probabilité que le lot contienne au moins deux stylo défectueux ?

Exercice 2 : (83 p.251, Repères, 1ère S, Hachette)

83. Loto Foot

Une grille de *Loto foot* comporte 15 matchs. Pour le match de l'équipe A contre l'équipe B, il y a trois choix possibles : l'équipe A gagne, l'équipe B gagne ou c'est un match nul. Le joueur doit faire des pronostics en cochant une case pour chaque match.
Un joueur remplit une grille au hasard.
Soit X la variable aléatoire représentant le nombre de bonnes réponses sur cette grille.

1. Quelles sont les valeurs que peut prendre X ?
2. Quelle loi suit X ? Donner ses paramètres.
3. Le plus gros lot est attribué au joueur ayant 15 bonnes réponses. Calculer $p(X=15)$ à 10^{-10} près.
4. D'autres gains sont attribués à ceux qui ont obtenu 12, 13 ou 14 bonnes réponses. Calculer ces probabilités.
5. En déduire $p(X \geqslant 12)$ à 10^{-6} près.
6. Combien de bonnes réponses peut-on espérer avoir en remplissant une grille au hasard ?

Exercice 3 : Le tir à l'arc. (Document Ressources 1ères)

A chaque tir, un archer atteint sa cible avec une probabilité égale à 0,7.
Combien de tirs doit-il effectuer pour que, avec une probabilité supérieure ou égale à 0,99, il atteigne la cible au moins deux fois ? Au moins une fois ?

Exercice 4 : (Enigme 2 p.223, Hyperbole, 1ère ES-L, Nathan)

Exercice 5 : contrôle de production. (Document ressources 1ères)

Une entreprise fabrique chaque jour 10 000 composants électroniques. Chaque composant présente un défaut avec la probabilité 0,002. Si le composant est repéré comme étant défectueux, il est détruit par l'entreprise, et chaque composant détruit fait perdre 1€ à l'entreprise.

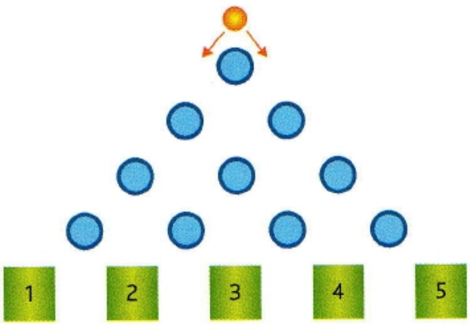
1. Les composants sont contrôlés un à un, et chaque contrôle coûte 0,1€. Quel est le coût moyen journalier pour l'entreprise (contrôles et destruction des composants défectueux) ?
2. Les composants sont regroupés par lots de 10, et on effectue un unique contrôle automatique de chaque lot, qui coûte lui aussi 0,1€. A l'issue de ce contrôle, le lot est accepté si tous les composants sont sains, et globalement détruit si l'un au moins des 10 composants présente un défaut. Quel est le coût moyen journalier pour l'entreprise avec ce nouveau dispositif (contrôles et destruction des composants défectueux) ?

Exercice 6 : Sondages.

Dans la population française, le pourcentage de femmes est d'environ 51,4% (source de 2005).
On tire au hasard un échantillon de 100 personnes dans cette population.
Quelle est la probabilité que l'échantillon contienne au moins 50 femmes ?

Exercice 7 (89 p.253, 1ère S, repères, Hachette)

89.



Une boule est lancée en haut d'une pyramide. À chaque obstacle, il y a une chance sur deux pour qu'elle se dirige à droite ou à gauche.

Soit X la variable aléatoire correspondant à la case où la boule tombe à la fin de son parcours.

Déterminer la loi de X.

I. 2. Fluctuations et prise de décision

Exercice 8 (49 p.166, Repères, 1ère S, Hachette)

49. Un cas particulier de l'inégalité de *Bienaymé-Tchébitchev*

Soit n un entier naturel non nul.

Soit $x_1, x_2, ..., x_n$ les valeurs d'une série statistique.

On note $\overline{x}$ sa moyenne et σ son écart type.

Soit k le nombre de valeurs de la série statistique vérifiant $|x_i - \overline{x}| < 2\sigma$. On suppose $\sigma > 0$.

1. Montrer que $\sum_{i=1}^{n} (x_i - \overline{x})^2 \geqslant 4(n-k)\sigma^2$.

2. En déduire que $k \geqslant \frac{3}{4} n$.

3. Recopier et compléter la phrase suivante :
«Au moins ....% des valeurs d'une série statistique appartiennent à l'intervalle $[\overline{x} - 2\sigma ; \overline{x} + 2\sigma]$.

4. En vous inspirant des questions précédentes, montrer qu'au moins 88 % des valeurs d'une série statistique appartiennent à l'intervalle $[\overline{x} - 3\sigma ; \overline{x} + 3\sigma]$.

Exercice 9 (Document ressources 1ères)

Monsieur Z, chef du gouvernement d'un pays lointain, affirme que 52% des électeurs lui font confiance.
On réalise un sondage dans cette population en interrogeant 100 électeurs au hasard (la population est supposée suffisamment grande pour considérer qu'il s'agit de tirages avec remise) et on souhaite savoir à partir de quelles fréquences observées sur notre échantillon, au seuil de 95%, on peut mettre en doute le pourcentage annoncé par Monsieur Z, dans un sens, ou dans l'autre.

k	$P(X \leq k)$
40	0,0106
41	0,0177
42	0,0285
43	0,0444
…	…
61	0,9719
62	0,9826

- On fait l'hypothèse que Monsieur Z dit vraie et que la proportion p des électeurs qui lui font confiance est égale à 0,52. Quelle est la loi de la variable aléatoire X, correspondant au nombre d'électeurs lui faisant confiance ?
- On donne ci-contre un extrait de la table des probabilités cumulées $P(X \leq k)$ où X suit la loi binomiale de paramètres $n=100$ et $p=0{,}52$.
  Déterminer a et b tels que :
  1. a est le plus petit entier tel que $P(X \leq a) > 0{,}025$;
  2. b est le plus petit entier tel que $P(X \leq b) \geq 0{,}975$.
- Comparer l'intervalle de fluctuation au seuil de 95%, $\left[\frac{a}{n};\frac{b}{n}\right]$, ainsi obtenu grâce à la loi binomiale, avec l'intervalle $\left[p-\frac{1}{\sqrt{n}};\, p+\frac{1}{\sqrt{n}}\right]$.
- Sur les 100 électeurs interrogés au hasard, 43 déclarent avoir confiance en Monsieur Z. Peut-on considérer au seuil de 95%, l'affirmation de Monsieur Z comme exacte ?

Exercice 10 (Document ressources 1ères)

Deux entreprises recrutent leur personnel dans un vivier comportant autant d'hommes que de femmes.
Voici la répartition entre hommes et femmes dans ces deux entreprises :

	Hommes	Femmes	Total
Entreprise A	57	43	100
Entreprise B	1350 (54%)	1150 (46 %)	2500

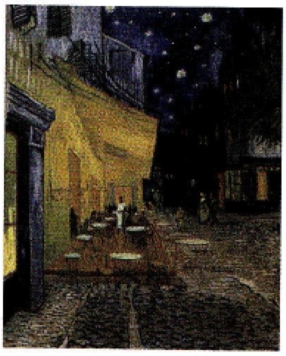
Peut-on suspecter l'une des deux de ne pas respecter la parité hommes-femmes à l'embauche ?

Exercice 11 (32 p.239, Hyperbole, 1ère ES-L, Nathan)

32 Art

On estime que 9 % des individus sont gauchers. Le tableau ci-dessous répertorie quelques artistes célèbres.

Michel Ange	Gaucher	XV^e-XVI^e s
Leonard de Vinci	Gaucher	XV^e-XVI^e s
Raphaël	Gaucher	XV^e-XVI^e s
Édouard Manet	Droitier	XIX^e s
Vincent Van Gogh	Droitier	XIX^e s
Claude Monet	Droitier	XIX^e-XX^e s
Pablo Picasso	Droitier	XIX^e-XX^e s
Salvadore Dali	Droitier	XX^e s
Yves Klein	Droitier	XX^e s

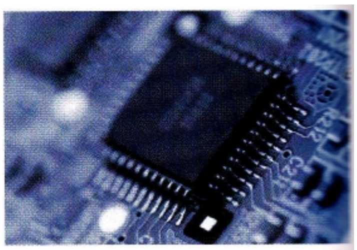


Terrasse de café le soir, Vincent Van Gogh.

La fréquence des peintres gauchers de ce tableau est-elle anormalement élevée ? Justifier votre raisonnement à l'aide d'un intervalle de fluctuation.

Exercice 12 (97 p254, repères, Hachette, 1ère S)

97. Une machine fabrique des processeurs. On sait que la probabilité d'obtenir un processeur défectueux est $p = 0{,}06$.



On contrôle des lots de 300 processeurs. Soit X la variable aléatoire représentant le nombre de processeurs défectueux sur ce lot.
On admet que X suit une loi binomiale de paramètres 300 et 0,06.

1. Déterminer la valeur du plus petit entier a tel que $p(X \leqslant a) > 0{,}025$.

2. Déterminer la valeur du plus petit entier b tel que $p(X \leqslant b) \geqslant 0{,}975$.

3. En déduite l'intervalle de fluctualion à 95 % d'une fréquence correspondant à la réalisation, sur un échantillon aléatoire de taille 300, de la variable aléatoire X.

4. Le contrôle de la machine A donne 23 processeurs défectueux ; le contrôle de la machine B donne 28 processeurs défectueux.
Que peut-on en conclure ?

Exercice 13 (Document ressources 1ères)

Une petite ville des Etats-Unis, Woburn, a connu 7 cas de leucémie parmi les 5969 garçons de moins de 15 ans sur la période 1969-1979. La fréquence des leucémies pour cette tranche d'âge aux Etats-Unis est égale à 0,00052. (Source : *Massachussetts Department of Public Health*).
Les autorités concluent qu'il n'y a rien d'étrange dans cette ville. Qu'en pensez-vous ?

Exercice 14 : Surbooking

Anticipant le fait que certains passagers ne se présentent pas le jour du vol, une compagnie aérienne souhaite vendre un nombre *n* de billets supérieur à la capacité d'embarquement de ses appareils.
Un cabinet d'experts lui affirme que la probabilité qu'un passager ne se présente pas le jour du vol est égale à $p=0,1$.
On note *X* la variable aléatoire qui représente le nombre de passagers se présentant le jour du vol.
a. Si *n* billets ont été vendus, quelle loi suit *X* ?
b. Donner un intervalle de fluctuations asymptotique à 95% de la fréquence *F=X/n*. En déduire un intervalle de fluctuations asymptotique à 95% de *X*, que l'on notera *IF*.
c. Les avions de la compagnie ont une capacité de 300 passagers. Afin de limiter le risque que trop de passagers se présentent le jour du vol, la compagnie propose de choisir pour *n* l'entier le plus grand tel que $IF\subset[0;300]$. Que vaut-il? On le note n_1 .
d. Pour ce choix $n=n_1$, quelle est la probabilité que le nombre de passagers se présentant le jour du vol dépasse la capacité de l'avion?
e. La compagnie considère qu'elle peut prendre un peu plus de risques. Elle souhaite choisir pour *n* l'entier le plus grand vérifiant $P(X>300)\leq 0,05$. On le note n_2 . Ecrire un algorithme permettant de trouver n_2 . Que vaut-il?
f. Pour le choix $n=n_2$, que vaut la probabilité que le nombre de passagers se présentant le jour du vol dépasse la capacité de l'avion?
g. Après quelques semaines de mise en pratique de sa politique de surbooking avec $n=n_2$, la compagnie fait un premier bilan : sur 200 vols effectués, un seul a dû géré l'arrivée d'un passager de trop. Cette observation est-elle en accord avec les choix effectués ci-dessus?

Partie II : Lois Continues.

II. 1. Calculs de probabilités

Exercice 15 (62 p.355, repères TS, Hachette)

1. Déterminer le réel k tel que, la fonction $f(x)=k|x|$ soit une densité de probabilité sur $[-3;1]$.
2. Soit X une variable aléatoire suivant une loi de densité f. Calculer $P(0\leq X\leq 1)$.

Exercice 16 (10 p.418, math'x, TS, Didier)

10 Visite expresso

Paolo vient tous les matins entre 7 heures et 7 heures 45 chez Lisa pour prendre un expresso.
Il peut débarquer à tout instant dans cette plage horaire avec les mêmes chances.

1. Proposer une loi de probabilité pour la variable aléatoire X modélisant l'heure d'arrivée de Paolo.

2. Calculer la probabilité que Paolo sonne chez Lisa

a. après 7 heures 30 **b.** avant 7 heures 10

c. entre 7 h 20 et 7 h 22 **d.** à 7 heures pile

Exercice 17 (application p.339, repères TS, Hachette)

Soit X une variable aléatoire suivant une loi exponentielle de paramètre $\frac{1}{2}$.

1. **Donner la densité de X.**
2. **Calculer $P(X \leqslant 4)$, puis en donner une valeur approchée à 0,01 près.**
   Vérifier ce résultat en utilisant le calcul intégral de la calculatrice.
3. **En déduire $P(X > 4)$.**
4. **Sachant que $(X \geqslant 2)$, calculer $P(X \geqslant 6)$.**

Exercice 18 (Application cours p.375, repères TS, Hachette)

TICE Calculer une probabilité pour la loi $\mathcal{N}(0 ; 1)$

Soit X une variable aléatoire suivant la loi $\mathcal{N}(0 ; 1)$.
En utilisant une calculatrice ou un logiciel, calculer à 10^{-4} près :

1. $P(-1 \leqslant X \leqslant 2)$; 2. $P(X \leqslant 2)$; 3. $P(X > 0,5)$.

Exercice 19 (63 p.398 , repères, TS, Hachette)

63. Soit Z une variable aléatoire suivant la loi $\mathcal{N}(0\,;1)$.
On donne $P(Z < 1) = 0{,}84$. Déterminer sans calculatrice les probabilités suivantes :

1. $P(X < 10)$ pour X suivant la loi $\mathcal{N}(8\,;4)$.
2. $P(X \geqslant 0)$ pour X suivant la loi $\mathcal{N}(-5\,;25)$.
3. $P(X < 0)$ pour X suivant la loi $\mathcal{N}(5\,;25)$.
4. $P(1 < X < 5)$ pour X suivant la loi $\mathcal{N}(5\,;16)$.

Exercice 20 (14 p.418, Math'x, TS, Didier)

Une variable aléatoire Y suit une loi exponentielle de paramètre $\lambda=2$.

1. Calculer t tel que $P(Y>t)=0{,}95$.
2. Calculer t' tel que $P(Y\leq t')=P(Y\geq t')$. Comment s'appelle le nombre t' ?

Exercice 21 (29 p.408, HyperboleTS, Nathan)

29 T est une variable aléatoire qui suit la loi normale centrée réduite et u est le nombre réel tel que :

$$P(-u \leqslant T \leqslant u) = 0{,}3$$

a) En déduire $P(0 \leqslant T \leqslant u)$, puis $P(T \leqslant u)$.

b) Utiliser la valeur de $P(T \leqslant u)$ pour déterminer, à l'aide de la calculatrice, l'arrondi au millième de u.

Exercice 22 (Document ressources Terminales)

4. ***Masse d'alerte pour cartes de contrôle***

Une coopérative produit du beurre en microplaquettes de 12,5g pour des collectivités et des chaînes hôtelières. Les microplaquettes sont conditionnées dans des boîtes de 40.

La masse des microplaquettes peut être modélisée par une variable aléatoire suivant une loi normale d'espérance $\mu = 12{,}5$ et de variance $\sigma^2 = 0{,}2^2$ et on admet que la variable aléatoire X égale à la masse d'une boîte de 40 microplaquettes suit alors une loi normale d'espérance $\mu = 500$ et de variance $\sigma^2 = 1{,}6$ (les notions relatives à la variance d'une somme de variables ne sont pas au programme, quelques notions sont abordées en annexe 2).

La boîte est jugée conforme si sa masse est comprise entre 496,2 g et 503,8 g (soit environ $500 \pm 3\sigma$).

1. Calculer la probabilité qu'une boîte prélevée aléatoirement en fin de chaîne de conditionnement soit non conforme. (*Réponse :* 0,003 *à* 10^{-3} *près)*
2. Pour contrôler le réglage de la machine, on détermine des poids d'alerte $\mu - h$ et $\mu + h$ tels que $P(\mu - h < X < \mu + h) = 0{,}99$. Ces poids d'alerte sont inscrits sur une carte de contrôle et correspondent à une marge de sécurité en lien avec des normes de conformité.

   Calculer les poids d'alerte.

II. 2. Modélisation

Exercice 23 (16 p.418, Math'x TS, Didier)

On considère que la durée de vie, en années, d'un élément radioactif est une variable aléatoire D qui suit une loi exponentielle de paramètre λ .

On appelle demi-vie de cet élément radioactif, le réel *t* tel que $P(D \leq t) = 0{,}5$.

a. Démontrer que : $t = \dfrac{\ln 2}{\lambda}$.

b. La demi-vie du Césium 137 est de 30 années.
Calculer la probabilité que la durée de vie d'un élément radioactif Césium 137 dépasse 50 ans.

Exercice 24 (49 p.387, Hyperbole TS, Nathan)

49 Une entreprise est spécialisée dans la location de voitures de luxe. La durée de vie de ses véhicules peut être modélisée par une loi exponentielle. Une étude statistique a montré que la durée de vie moyenne d'un véhicule de cette entreprise est de 2 000 jours.

a) Déterminer le paramètre de cette loi.

b) Quelle est la probabilité qu'un véhicule de location choisi au hasard dans cette entreprise fonctionne encore au bout de 3 ans ? Arrondir au centième.

Exercice 25 (51 p.410, Hyperbole TS, Nathan)

51 Une machine permet l'emballage de paquets de farine. On obtient des masses distribuées normalement autour d'une valeur *m*, et d'écart-type 0,002 5 kg.
Sur quelle valeur *m* doit-on régler la machine pour que 90 % des paquets aient une masse supérieure à celle indiquée sur le paquet, qui est 1 kg ?

Exercice 26 (53 p.410, Hyperbole TS, Nathan)

53 Dans une usine, une machine produit des barres de métal. On définit la variable aléatoire X qui à chaque barre associe sa longueur et on admet que la variable aléatoire X suit la loi normale d'espérance $\mu = 92{,}5$ et d'écart-type σ.

Une barre de la production est mise au rebut si sa longueur est inférieure à 92,2 cm ou supérieure à 92,8 cm.

1. On suppose dans cette question que $\sigma = 0{,}20$.

a) Calculer la probabilité qu'une barre extraite au hasard de la production de la machine soit mise au rebut.

b) Déterminer le nombre réel *a* tel que la probabilité que la variable aléatoire X prenne des valeurs comprises entre $92{,}5 - a$ et $92{,}5 + a$ soit égale à 0,90.

2. Quelle valeur faudrait-il donner à σ pour que la probabilité de mise au rebut soit égale à 0,01 ?

Exercice 27 (82 p.402, repères TS, Hachette)

82. Un laboratoire pharmaceutique souhaite produire un nouveau médicament. Il faut pour cela ajuster le dosage de la substance active dans chaque comprimé. Plusieurs techniques industrielles sont envisageables : toutes fournissent un dosage aléatoire X, plus ou moins précis, suivant une loi $\mathcal{N}(\mu ; \sigma^2)$, où $\mu \geqslant 0$ et $\sigma > 0$. À chaque technique industrielle est associée un unique σ reflétant la précision du dosage induit. De plus, quelle que soit la technique choisie, le dosage moyen peut être réglé au travers du paramètre μ.

Si la teneur en substance active est trop importante, le médicament peut devenir dangereux. Le laboratoire considère ainsi que X doit être inférieur à 60 microgrammes avec une probabilité supérieure ou égale à 0,999. Inversement, si X est trop faible, le médicament n'est plus efficace ; le laboratoire souhaite donc que X soit supérieur à 40 microgrammes avec une probabilité supérieure ou égale à 0,95.

1. Justifier que les deux exigences du laboratoire se traduisent par le système :

$$\begin{cases} P(X \leqslant 40) \leqslant 0{,}05 \\ P(X \leqslant 60) \geqslant 0{,}999 \end{cases}.$$

2. Soit Z une variable aléatoire suivant la loi $\mathcal{N}(0 ; 1)$. Écrire le système précédent en fonction de Z.

3. En déduire que μ et σ doivent vérifier :

$$\begin{cases} \dfrac{40-\mu}{\sigma} \leqslant -1{,}64 \\ \dfrac{60-\mu}{\sigma} \geqslant 3{,}09 \end{cases}.$$

4. Supposons que l'on choisisse une technique de dosage industrielle particulière. Cela revient à fixer $\sigma > 0$. À quel intervalle, dépendant de σ, doit alors appartenir μ ?

5. En déduire la valeur maximale possible pour le choix de σ. On la notera σ_{max}.

6. Plus la technique industrielle est précise dans son dosage, plus elle est chère à mettre en œuvre. Le laboratoire désirant minimiser son coût de production, il a intérêt à choisir une technique industrielle associé au σ le plus grand possible, tout en permettant de respecter les exigences de dosage. Le laboratoire choisit donc la technique industrielle associée à σ_{max}. Quelle valeur du dosage moyen μ devra-t-il enfin choisir pour respecter ses exigences ?

Exercice 28 (Document ressources Terminales)

5. ***Réglage d'une machine d'embouteillage dans une coopérative***

Sur une chaîne d'embouteillage dans une brasserie, la quantité X (en cL) de liquide fournie par la machine pour remplir chaque bouteille de contenance 110 cL peut être modélisée par une variable aléatoire de loi normale de moyenne μ et d'écart-type $\sigma = 2$.

La législation impose qu'il y ait moins de 0,1% de bouteilles contenant moins d'un litre.

1. A quelle valeur de la moyenne μ doit-on régler la machine pour respecter cette législation ?
2. La contenance des bouteilles étant de 110 cL, quelle est alors la probabilité qu'une bouteille déborde lors du remplissage ?
3. Le directeur de la coopérative veut qu'il y ait moins de 1% de bouteilles qui débordent au risque de ne plus suivre la législation.
   a) Quelle est alors la valeur de μ ?
   b) Quelle est dans les conditions de la question a) la probabilité que la bouteille contienne moins d'un litre?
   c) Déterminer la valeur moyenne μ et l'écart-type maximal possible σ_{max} afin qu'il y ait moins de 0,1% de bouteilles de moins d'un litre ET moins de 1% de bouteilles qui débordent.

Exercice 29 (69 p.399, repères TS, Hachette)

69. R Commander Un fabricant de terreau distribue sa production en sacs de 50 litres. Afin d'évaluer la qualité de sa mise en sac, il mesure la quantité de terreau de 20 sacs. Le tableau suivant résume ce qu'il a relevé (en litres).

50,46	50,11	49,75	51,10	47,35
49,77	50,03	48,95	51,06	49,68
50,16	52,10	49,74	49,35	50,06
49,41	50,01	48,86	50,28	50,70

Rappel
R Commander se lance en tapant dans R :
library(Rcmdr)

1. Entrer ces valeurs dans la première colonne d'un tableau.
Dans R Commander *:* **Données/Nouveau jeu de données**

2. Calculer la moyenne et l'écart-type de cet échantillon.
Dans R Commander *:*
Statistiques/Résumés/Statistiques descriptives

Le producteur est satisfait de la moyenne relevée : elle lui paraît conforme aux 50 litres attendus. Par contre, il s'inquiète qu'un sac puisse contenir plus de 52 litres de terreau. Afin d'approfondir son étude il suppose que la quantité de terreau dans chaque sac fluctue selon la loi normale de moyenne $\mu = 50$ et d'écart-type σ.

3. Par quelle valeur peut-on approcher σ ?

On suppose dans la suite que $\sigma = 1$. Vérifions dans un premier temps l'hypothèse du producteur.

4. Créer une nouvelle colonne contenant l'échantillon centré et réduit, c'est-à-dire contenant les valeurs de $\frac{(X-\mu)}{\sigma}$, où X désigne la quantité de terreau relevé.
Dans R Commander *:* **Données/Gérer les variables dans le jeu de données actif/Calculer une nouvelle variable**

5. Tracer l'histogramme de l'échantillon centré et réduit.
Dans R Commander *:* **Graphes/Histogramme** *et choisir l'option* **Densité**.

L'histogramme semblant raisonnablement proche de la courbe de la densité de la loi $\mathcal{N}(0\,;1)$, le producteur valide son hypothèse : la quantité de terreau dans chaque sac suit une loi $\mathcal{N}(\mu\,;\sigma^2)$.

6. Selon cette hypothèse, quelle est la probabilité qu'un sac contienne plus de 52 litres de terreau ?

Dans R Commander *:* **Distributions/Distributions continues/Distribution normale/Probabilités normales**

7. Quelle devrait être la valeur de l'écart-type σ pour que cette probabilité n'excède pas 1 % ?
Dans R Commander *:* **Distributions/Distributions continues/Distribution normale/Quantiles normaux**

Aide
Centrer et réduire pour se ramener à la loi $\mathcal{N}(0\,;1)$.

Exercice 30 : modéliser d'un jeu de données réel

Ouvrir le fichier quelle_loi.txt avec RCommander.

1. Calculer la moyenne et l'écart-type de cet échantillon.
2. A l'aide d'une représentation graphique, est-il plus raisonnable de modéliser la répartition de ces valeurs selon une loi uniforme, une loi exponentielle ou une loi normale ?
3. Estimer le(s) paramètre(s) de la loi retenue précédemment.

II. 3. Retour à la loi Binomiale

a) Moivre-Laplace

Exercice 31 (TP p.418, repères TS, Hachette)

L'universalité de la loi normale

objectif

Vérifier le théorème de Moivre-Laplace et observer sa généralisation : le Théorème Limite Central.

1. Le théorème de Moivre-Laplace vu comme une loi des erreurs

On dispose de $n = 100$ appareils (numérotés de 1 à 100) pour mesurer une certaine quantité inconnue. Chaque appareil souffre d'une erreur de mesure aléatoire. On note E_i l'erreur de mesure de l'appareil numéro i.

Dans cette première partie, on suppose (de façon simpliste) que chaque erreur E_i ne peut prendre que les valeurs 0 ou 1 : E_i vaut 1 avec la probabilité 0,5 et 0 avec la probabilité 0,5.

On effectue les mesures de manière successive et indépendante par les 100 appareils et on s'intéresse à la somme des erreurs de mesure $X_n = \sum_{i=1}^{100} E_i$.

a. Justifier que X_n suit la loi binomiale $\mathcal{B}(100\,;0{,}5)$. Donner l'espérance $E(X_n)$ et l'écart-type $\sigma(X_n)$ de X_n.

b. Simuler dans la première ligne d'un tableau une réalisation des 100 erreurs de mesure $E_1, \ldots, E_{100}$ puis en calculer la somme X_n dans la dernière colonne.

Avec R Commander : il suffit d'utiliser le menu **Distributions/ Distributions discrètes/Distribution binomiale/ Echantillon d'une distribution binomiale**. Dans le menu qui apparaît on fixe les paramètres de la loi binomiale (ici une loi de Bernoulli) : **Nombre d'essais** à 1 et **Probabilité de succès** à 0,5. Le **Nombre d'échantillons** que l'on souhaite doit être fixé à 1 et le **Nombre d'observations** à 100. Pour calculer la somme X_n, il suffit de cocher **Somme des échantillons**. Le tableau créé peut être visualisé en cliquant sur le bouton Visualiser.

c. Simuler dans un nouveau tableau 1 000 réalisations des 100 erreurs de mesures $E_1, \ldots, E_{100}$ puis en déduire les 1 000 réalisations de leur somme X_n.

Avec R Commander : procéder comme en **b.** en fixant le **Nombre d'échantillons** à 1 000. Vérifier en cliquant sur Visualiser que le tableau créé contient en dernière colonne les 1 000 réalisations de X_n.

d. Ajouter au tableau une colonne, que l'on nommera « X_n *centré réduit* », valant $\dfrac{X_n - E(X_n)}{\sigma(X_n)}$.

Avec R Commander : aller dans **Données/Gérer les variables dans le jeu de données actif/Calculer une nouvelle variable**.

e. Tracer l'histogramme des 1 000 réalisations de « X_n *centré réduit* ».

Avec R Commander : aller dans **Graphes/Histogramme** et choisir l'option **Densité**.

f. Quel théorème le résultat précédent illustre-t-il ?

2. Vers la loi universelle des erreurs : le théorème limite central

On suppose dans cette partie que les erreurs E_i ne sont pas restreintes à valoir uniquement 0 ou 1. Reprendre les questions **c.**, **d.**, **e.** et **f.** dans les cas suivants :

(i) Chaque erreur fluctue selon une loi exponentielle $\mathcal{E}(2)$. On admet que $E(X_n) = 50$ et $\sigma(X_n) = 50$.

(ii) Chaque erreur fluctue selon n'importe quelle loi au choix proposée dans R Commander, mise à part la loi de Cauchy. On prendra alors pour $E(X_n)$ la moyenne des réalisations de X_n et pour $\sigma(X_n)$ leur écart-type.

Que remarque-t-on ?

Commentaire : ce TP illustre un résultat fondamental de la statistique, le théorème limite central, qui généralise le théorème de Moivre-Laplace : quelle que soit la façon dont fluctuent les erreurs, si elles sont nombreuses leur somme ou leur moyenne fluctue (presque) toujours comme une loi normale.

Exercice 32 (90 p.403, repères TS, Hachette)

90. Soit X une variable aléatoire suivant la loi binomiale $\mathcal{B}(1760\,;0{,}1)$.

1. Calculer à l'aide d'une calculatrice $P(X \geqslant 180)$.
2. Compléter la définition suivante en remplaçant les points d'interrogation par leur valeur :

$$P(X \geqslant 180) = \sum_{k=?}^{?} \binom{1760}{k} 0{,}1^k\, 0{,}9^?$$

3. Essayer de calculer le premier terme de cette somme à l'aide d'une calculatrice.

Pour calculer $P(X \geqslant 180)$, les calculatrices ne peuvent pas utiliser la définition précise rappelée en 2. : elles utilisent une approximation.

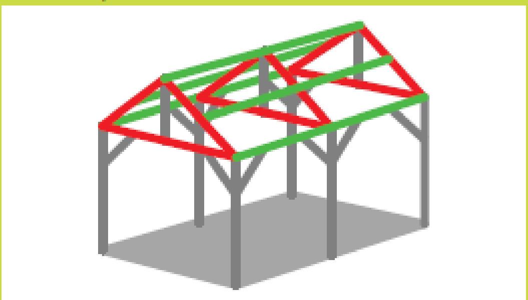
4. Justifier que les conditions pratiques d'utilisation du théorème de Moivre-Laplace sont réunies.
5. Calculer la probabilité précédente en utilisant l'approximation fournie par le théorème de Moivre-Laplace.
6. Le résultat est-il le même qu'en 1. ?

Commentaire : les calculatrices utilisent en fait une autre approximation, plus précise que celle fournie par le théorème de Moivre-Laplace, basée sur la *formule de Stirling*.

Exercice 33 (97 p.405, repères TS, Hachette)

Le saviez-vous ?

Les « fermes » (en rouge sur le schéma) sont les parties principales d'une charpente d'une toiture. Elles sont reliées entre elles par des « pannes » (en vert sur le schéma).



97. Certaines poutres nécessaires à la réalisation des fermes doivent avoir une section rectangulaire de dimensions très précises. On admet que la probabilité qu'une poutre, prise au hasard dans la production, soit conforme, est 0,9. Une commande nécessite 1 500 poutres. On veut évaluer la probabilité que le nombre de poutres non conformes soit d'au plus 160. On note X la variable aléatoire qui, à chaque lot de 1 500 poutres, associe le nombre de poutres non conformes.

1. Quelle est la loi suivie par X ? Que valent son espérance E(X) et son écart-type $\sigma(X)$?
2. En quoi le calcul de $P(X > 1\,340)$ est-il « difficile » ?
3. Pour ce calcul, on décide d'approcher la loi de $\dfrac{X - E(X)}{\sigma(X)}$ par la loi normale centrée réduite. Justifier ce choix.
4. En utilisant cette approximation, calculer une valeur approchée à 10^{-3} près de $P(X > 1\,340)$.

D'après BTS Agencement de l'environnement architectural, 2007.

b) Intervalle de fluctuations et prise de décision basés sur l'approximation de Moivre-Laplace.

Exercice 34 : « Monsieur Z » (suite)

Retour à l'exercice « Monsieur Z » (partie I).
Déterminer l'intervalle de fluctuation asymptotique issu de la loi normale et le comparer aux deux autres.

Exercice 35 : Surbooking (suite)

Retour à l'exercice sur le surbooking (partie I)
Déterminer le plus grand entier n_3 vérifiant $P(X>300)\leq 0{,}05$ en utilisant l'approximation normale et comparer n_1 n_2 et n_3 .

Exercice 36 (110 p.407 repères TS, Hachette)

110. Un fabricant de diodes électroluminescentes (LED) garantit que la probabilité p qu'une diode ne fonctionne pas vaut au plus 0,03. Pascal s'est fait livrer 5 000 diodes. On note X le nombre de diodes défectueuses parmi 5 000 diodes et $F=\dfrac{X}{5\,000}$ leur proportion.

1. Quelle est la loi suivie par X si $p=0{,}03$?
2. D'après l'approximation de Moivre-Laplace, dans quel intervalle fluctue F avec une probabilité de 0,95 ?
3. On en déduit que X fluctue à plus de 95 % dans l'intervalle $[127\ ;\ 174]$. Vérifier cette affirmation en calculant $P(127 \leqslant X \leqslant 174)$.

Pascal a constaté que 172 diodes ne fonctionnaient pas dans le lot de 5 000 diodes qu'il a commandé. Il trouve ce nombre trop élevé. Le fabricant lui explique que si ce nombre appartient à l'intervalle de fluctuations à 95 % de X, il n'y a pas de raison de considérer ce lot comme non conforme.

4. Suivant cette règle de décision, quelle est la probabilité qu'un lot ne soit pas conforme ?
5. Suivant cette règle de décision, le lot commandé par Pascal est-il non-conforme ?
6. Si le lot de Pascal n'avait contenu aucune diode défectueuse, aurait-il été considéré comme conforme ?

Pascal trouve la règle de décision absurde. Il propose une autre règle : si le lot contient moins de 170 diodes défectueuses, alors il est jugé conforme, sinon il n'est pas jugé conforme.

7. Suivant la règle de décision de Pascal, quelle est la probabilité qu'un lot ne soit pas conforme ?
8. Un lot ne contenant aucune diode défectueuse serait-il jugé conforme ?
9. Quelle règle de décision vous paraît la plus adaptée au problème : celle du fabricant ou celle de Pascal ?
10. Selon sa règle de décision, le lot reçu par Pascal est-il conforme ?

Exercice 37 (Document ressources Terminales)

Dans les exemples qui suivent, les tirages sont effectués sans remise. Toutefois, la taille des échantillons considérés étant faible par rapport à la taille de la population totale, on apparente les tirages à des tirages avec remise, correspondant alors à un schéma de Bernoulli et permettant d'appliquer les résultats théoriques précédents.

1. ***Prise de décision***

On admet que dans la population d'enfants de 11 à 14 ans d'un département français le pourcentage d'enfants ayant déjà eu une crise d'asthme dans leur vie est de 13%.
Un médecin d'une ville de ce département est surpris du nombre important d'enfants le consultant ayant des crises d'asthme et en informe les services sanitaires. Ceux–ci décident d'entreprendre une étude et d'évaluer la proportion d'enfants de 11 à 14 ans ayant déjà eu des crises d'asthme.
Ils sélectionnent de manière aléatoire 100 jeunes de 11 à 14 ans de la ville.

La règle de décision prise est la suivante : si la proportion observée est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 95% alors une investigation plus complète sera mise en place afin de rechercher les facteurs de risque pouvant expliquer cette proportion élevée.

1) Déterminer l'intervalle de fluctuation asymptotique au seuil de 95% de la proportion de jeunes de 11 à 14 ans ayant eu une crise d'asthme dans un échantillon de taille 100. (*solution* : [0,06 ; 0,20])

2) L'étude réalisée auprès des 100 personnes a dénombré 19 jeunes ayant déjà eu des crises d'asthme. Que pouvez-vous conclure ?

Solution : la valeur 0,19 est à l'intérieur de l'intervalle de fluctuation asymptotique au seuil de 95%, On en conclut que la règle de décision choisie ne prévoit pas de réaliser une enquête supplémentaire.

3) Le médecin n'est pas convaincu par cette conclusion et déclare que le nombre de personnes interrogées était insuffisant pour mettre en évidence qu'il y avait plus de jeunes ayant eu des crises d'asthme que dans le reste du département.
Combien faudrait-il prendre de sujets pour qu'une proportion observée de 19% soit en dehors de l'intervalle de fluctuation asymptotique ?

Exercice 38 (111 p.408 repères TS, Hachette)

111. M. Frit, gérant d'une boutique de forfaits téléphoniques, vient d'acheter un lot de 100 cartes SIM vierges à son distributeur Dominique. Certaines cartes peuvent présenter des défauts et s'avérer inutilisables. Mais Dominique garantit que la probabilité p qu'une carte SIM soit endommagée vaut au maximum 0,05.

M. Frit constate que 9 cartes SIM dans le lot de 100 étaient endommagées, soit 9 %. Il a l'impression d'avoir été trompé par Dominique sur la qualité des cartes et s'en plaint. Dominique lui affirme que trouver 9 cartes endommagées sur 100 est tout a fait conforme aux 5 % annoncés. Pour le démontrer, il lui propose de faire un test pour décider entre l'hypothèse « $p = 0{,}05$ » (ce qu'il a prétendu) et l'hypothèse « $p > 0{,}05$ » (ce que pense M. Frit). La règle sera la suivante : si $p = 0{,}05$, la proportion de cartes SIM endommagées dans un lot de 100 devrait fluctuer à 95% dans un certain intervalle I. Donc si cet intervalle I contient les 9 % observés par M. Frit, il n'y a aucune raison de remettre en cause l'hypothèse $p = 0{,}05$.

On notera X le nombre de cartes endommagées dans un lot de 100, et $F = \frac{X}{100}$ leur proportion.

1. Montrer que l'intervalle de fluctuations dont parle Dominique est $I \approx [0{,}007 \ ; 0{,}093]$.

2. Quelle est la décision du test ?

M. Frit n'est pas satisfait et demande à Dominique quelle est la fiabilité de son test. Dominique lui explique que la probabilité de se tromper correspond à la probabilité que F n'appartienne pas à I alors que p vaut 0,05. Il lui assure qu'elle est très faible.

3. Montrer qu'en suivant la définition de Dominique, la probabilité de se tromper vaut environ 5 %.

M. Frit n'abandonne pas : il se demande quelle serait la probabilité que F appartienne à I si p valait 0,09 et non 0,05.

4. En supposant $p = 0{,}09$, quelle loi suit X ?

5. Montrer que $p(F \in I) = P(7 \leqslant X \leqslant 93)$ et calculer cette probabilité lorsque $p = 0{,}09$.

6. Le test vous paraît-il à l'avantage de Dominique ou de M. Frit ?

M. Frit contacte son collègue M. Hésséfer et lui raconte sa mésaventure. M. Hésséfer avait lui aussi commandé un lot de 100 cartes SIM à Dominique. Il avait relevé 8 cartes défectueuses, s'en était plaint à Dominique, mais avait été convaincu par le résultat du test.

7. Effectuer le test pour le lot de M. Hésséfer. Quelle est la conclusion ?

La conclusion n'étonne pas M. Frit. Il propose à M. Hésséfer de mutualiser leurs données.

8. Soit F' la proportion de cartes défectueuses dans les deux lots. Dans quel intervalle I' fluctue F' à 95 % si $p = 0{,}05$?

9. Effectuer le test de Dominique en utilisant toutes les données de M. Frit et M. Hésséfer. Quelle est la conclusion ?

c) Intervalle de confiance asymptotique pour l'estimation de p

Exercice 39 (5 p.391, repères TS, Hachette)

Exercice résolu 5 TICE Simuler et analyser des sondages

énoncé Dans une population de 100 000 personnes, une proportion p valant 35 % de la population achètent régulièrement leurs chaussures sur internet. Soit X la variable aléatoire égale au nombre d'individus, dans un échantillon de 1 000 personnes tirées au hasard, achetant régulièrement leurs chaussures sur internet. Soit $F = \frac{X}{1000}$ la fréquence de ces individus dans l'échantillon. On admet que, compte tenu du nombre de personnes sondées par rapport à la taille de la population, on peut assimiler le tirage à un tirage avec remise.

1. Quelle est la loi de X ?
2. Dans quel intervalle fluctue F avec une probabilité de 0,95 ?

On suppose qu'on ne connaît pas la proportion p et on désire l'estimer à partir d'un sondage sur 1 000 personnes. Nous allons créer sous *Xcas* une population de 100 000 personnes dont 35 % prennent la valeur 1 (ils achètent leurs chaussures régulièrement sur internet) et les autres prennent la valeur 0 (ils n'achètent pas régulièrement leurs chaussures sur internet). Nous allons ensuite simuler un sondage sur 1 000 personnes pour estimer la proportion p.

3. Créer un vecteur pop contenant 35 000 fois la valeur 1 et 65 000 fois la valeur 0. Pour cela, on pourra taper : pop:=concat(seq(1,k,1,35000),seq(0,k,1,65000)) ;
4. Tirer un échantillon ech de 1 000 personnes avec la commande ech:=rand(1000,pop) ;
5. Que vaut la fréquence des 1 dans l'échantillon tiré ? (utiliser la fonction mean).
6. Donner un intervalle de confiance au niveau 95 % pour l'estimation de p, basé sur l'échantillon précédent.

Évaluons à présent la fluctuation de cette estimation en simulant plusieurs sondages.

7. Que contiennent les trois vecteurs frequences, ICinf et ICsup créés par le programme suivant :

```
ech:=rand(1000,pop) ; frequences:=mean(ech) ;
ICinf:=mean(ech)–1/sqrt(1000) ;    ICsup:=mean(ech)+1/sqrt(1000) ;
for(k:=1;k<100;k++){
    ech:=rand(1000,b) ;
    freq:=mean(ech) ; frequences:=frequences,freq ;
    ICinf:=ICinf,freq-1/sqrt(1000) ; ICsup:=ICup,freq+1/sqrt(1000);}
```

8. Lancer le programme précédent.
   Combien d'intervalles de confiance à 95 % contiennent la vraie valeur de p ? *(on pourra utiliser les fonctions count_inf et count_sup)*
9. Représenter les intervalles de confiance à l'aide du programme suivant :

```
line(x=0.35) ;
for(k:=0;k<100;k++){
    if(ICinf[k]< 0.35 && 0.35<ICsup[k]){
        segment(point(ICinf[k],k+1),point(ICsup[k],k+1),couleur=vert);}
    else{segment(point(ICinf[k],k+1),point(ICsup[k],k+1), couleur=rouge);}
       point(frequences[k],k+1, couleur=bleu);}
```

Exercice 40 (112 p.408, repères TS, Hachette)

112. corrigé Soit X une variable aléatoire suivant une loi $\mathcal{B}(267\,;p)$, où $p \in]0\,;\,1[$ est inconnu.

1. À quel intervalle doit appartenir p pour qu'il soit raisonnable d'utiliser l'approximation normale fournie par le théorème de Moivre-Laplace ?

2. Soit $F = \frac{X}{267}$. On observe à partir d'une réalisation de X, que $F = 0{,}3$. À partir de cette observation, donner un intervalle de confiance au niveau asymptotique 95 % pour estimer p.

3. Cet intervalle de confiance repose sur l'approximation normale issue du théorème de Moivre-Laplace. D'après le résultat de la question 2. et les conditions établies dans la question 1., peut-on accepter *a posteriori* d'avoir utilisé cette approximation ?

Exercice 41 (Document ressources Terminales)

1. *Diagnostic de la jaunisse*

Un test de diagnostic rapide effectué sur des sujets ictériques (coloration jaune de la peau, des muqueuses -couche de cellules de protection recouvrant les organes creux en contact avec l'extérieur- et du blanc de l'œil –sclérotique-) doit permettre d'estimer si l'ictère est d'origine virale ou non, sans avoir besoin de faire des analyses longues et compliquées. Cependant il est important de pouvoir s'assurer que ce test est de bonne qualité c'est-à-dire qu'il doit pouvoir indiquer correctement si l'ictère est viral ou non. Il doit être capable d'identifier correctement le type d'ictère : il est positif chez les sujets dont l'ictère est viral et négatif sinon.

Une étude est effectuée sur 100 personnes ayant un ictère viral et 100 personnes ayant un ictère d'origine non virale.

Les résultats obtenus sont présentés dans le tableau ci-dessous

	Hépatite virale	Ictère d'origine non virale
Test positif	85	20
Test négatif	15	80

a) Déterminer la proportion de sujets ayant un test positif parmi ceux ayant un ictère viral.

b) Déterminer un intervalle de confiance à 95% de la proportion de tests positifs lorsque l'ictère est viral. Cette proportion est appelée sensibilité du test diagnostic, c'est-à-dire la probabilité qu'une personne ayant un ictère viral réagisse au test. Un test diagnostic sera d'autant meilleur que la sensibilité est importante. (*réponse* : [0,75 ; 0,95]).

c) Déterminer la proportion de sujets ayant un test négatif parmi celles ayant un ictère non viral.

d) Déterminer un intervalle de confiance à 95% de la proportion de tests négatifs lorsque l'ictère est non viral. Cette proportion est appelée spécificité du test diagnostic, c'est-à-dire la probabilité qu'une personne ayant un ictère non viral ne réagisse pas au test. Un test diagnostic sera d'autant meilleur que la spécificité est importante.(*réponse* : [0,7 ; 0,9]).

Exercice 42 (122 p.411, repères TS, Hachette)

122. Deux candidats s'affrontent lors du second tour de l'élection présidentielle. Un institut de sondage souhaite effectuer une enquête auprès d'un échantillon de n personnes tirées aléatoirement afin de connaître les intentions de vote. Étant donné la taille de la population par rapport à n, ce tirage peut être considéré comme un tirage avec remise. L'institut de sondage pense que le scrutin sera serré et souhaite effectuer un sondage suffisamment précis pour pouvoir prédire le vainqueur. Combien de personnes n l'institut doit-il interroger pour que les intervalles de confiance au niveau asymptotique 95 % ait une longueur inférieure à 0,02 ?

Exercice 43 (Document ressources Terminales)

3 *Comparaison du taux de germination de semences de tomates de l'année avec celles de l'année précédente.*

Un maraîcher achète un lot de semences de tomates pour produire ses plants de tomate. Il lui reste des semences de l'année passée, dont il doit contrôler le taux de germination pour pouvoir les utiliser avec les autres. En effet, des taux de germination trop différents provoquent des trous dans les plates bandes de production, ce qui génère un coup de manutention plus élevé (il faut enlever les pots non germés avant de les conditionner). Il faut donc comparer les taux de germination des semences des deux années.

Une stratégie (il en existe d'autres, hors programme, mais qui peuvent faire l'objet d'une recherche) consiste à calculer et à comparer les intervalles de confiance des taux de germination (qui sont des proportions) des plants de l'année et de l'année précédente. Si les deux intervalles ne se recoupent pas, on peut conclure à une différence de taux de germination entre les semences des deux origines[16]. Il faudra alors les semer séparément.

Pour faire cette comparaison, le maraîcher prélève, aléatoirement dans les semences de l'année, un échantillon de 200 graines qu'il met à germer. Il constate que 185 graines germent.

Il prélève ensuite, aléatoirement dans les semences de l'année précédente, un échantillon de 200 graines qu'il met à germer. Il constate que 150 graines germent.

1. Déterminer un intervalle de confiance, au niveau de confiance de 95%, du taux de germination p_a du lot de semences de l'année.

2. Déterminer (par la même méthode qu'à la question a)) un intervalle de confiance au niveau 95%, du taux de germination p_b du lot de semences de l'année précédente.

3. Conclure.

Exercice 44 (123 p.411, repères TS, Hachette)

123. Un établissement hospitalier pratique l'implant de prothèses en silicone. Quelques rares cas de ruptures de la prothèse sont régulièrement à déplorer. Mais l'établissement s'inquiète du taux de ruptures anormal des prothèses de la marque TIT. En effet sur 81 prothèses TIT implantées, 6 cas de ruptures ont été déclarées, contre 14 cas de ruptures sur 1 303 implants pour l'ensemble des autres marques de prothèses.

1. Calculer le taux de ruptures des prothèses TIT et le taux de ruptures des prothèses des autres marques constatés dans cet établissement.

2. Proposer un intervalle de confiance au niveau asymptotique 95 % pour la probabilité de ruptures des prothèses de la marque TIT.

3. Même question pour la probabilité de ruptures des prothèses des autres marques.

4. D'après les deux questions précédentes, peut-on affirmer qu'il y a une probabilité significativement plus importante qu'une prothèse de la marque TIT se rompe ?

5. On suppose qu'il y aura toujours environ 16 fois plus d'implants de prothèses des autres marques que de la marque TIT. On note n le nombre d'implants de prothèses de la marque TIT totalisés à un instant donné dans l'établissement. Il y a donc $16n$ implants de prothèses des autres marques au même moment. Combien d'implants n faudra-t-il atteindre pour qu'un taux de 7,4 % de ruptures de prothèses TIT puisse être considéré significativement supérieur au taux de ruptures 1,1 % chez les autres marques ?

6. Par principe de précaution, le chef de l'établissement hospitalier ne souhaite pas attendre d'avoir atteint ce nombre et il alerte les autorités sanitaires. Ces dernières lancent immédiatement une enquête et recueillent les informations suivantes auprès de plusieurs établissements hospitaliers : sur 2 017 implants de prothèses TIT, 95 ruptures ont été déplorées alors que sur 29 814 implants de prothèses d'autres marques, 309 ruptures sont survenues. Sur la base de ces données, peut-on affirmer que la probabilité de ruptures des prothèses TIT est supérieure à celle des autres marques ?

Exercice 45 : autour de la représentativité d'un échantillon (Document ressources Terminales)

On souhaite estimer la prévalence du surpoids dans une ville V, c'est-à-dire la proportion de personnes ayant une masse trop importante par rapport à leur taille. Pour cela 460 personnes ont été sélectionnées de manière aléatoire à partir de la liste des logements connue par la municipalité, c'est-à-dire que le fait d'avoir été sélectionné pour participer à l'étude est uniquement dû au hasard. On admet que cette procédure permet d'assimiler la sélection des personnes interrogées à un schéma de Bernoulli. Un enquêteur s'est déplacé au sein de chaque logement après avoir convenu d'un rendez-vous afin de recueillir les informations nécessaires à l'enquête.

1° Dans un premier temps, l'enquêteur va s'assurer que l'échantillon est représentatif de la population qu'on étudie sur des informations qu'on peut vérifier et qui sont en lien avec le critère étudié. Dans le cas présent on peut connaître par exemple la proportion d'hommes et de femmes dans la population de la ville, ainsi que la répartition selon l'âge en demandant à la municipalité qui se référera aux informations du recensement. Parallèlement on peut comptabiliser le nombre d'hommes et de femmes dans l'échantillon ainsi que la répartition selon l'âge.

	Homme	Femme	Total
Echantillon	200	260	460

	< 60 ans	> 60 ans	Total
Echantillon	352	108	460

On sait que, dans la population, il y a 46% d'hommes et 20% de personnes de plus de 60 ans.

a) Déterminer l'intervalle de fluctuation asymptotique au seuil 0,95 de la variable aléatoire « proportion de femmes» dans un échantillon aléatoire de taille 460 sélectionné au sein de la population de cette ville.
b) Calculer la proportion de femmes dans l'échantillon et vérifier si cette valeur appartient à l'intervalle de fluctuation.
c) Déterminer l'intervalle de fluctuation asymptotique au seuil 0,95 de la variable aléatoire « proportion de personnes âgées de plus de 60 ans » dans un échantillon aléatoire de taille 460 sélectionné au sein de la population de cette ville.
d) Calculer la proportion de personnes de plus de 60 ans dans l'échantillon et vérifier si cette valeur appartient à l'intervalle de fluctuation.
e) Si pour chacune des variables, genre et âge, l'intervalle de fluctuation asymptotique au seuil de 95% contient la valeur de l'échantillon on considère que l'échantillon est représentatif de la population pour cette information. Quelle est donc la conclusion pour le cas étudié ici ?

2° Dans le cas de l'étude présentée ici, on souhaite estimer la proportion de personnes en surpoids ; pour cela il est tout d'abord important de définir le surpoids. La définition du surpoids donnée par l'OMS (Organisation Mondiale de la Santé) est la suivante : une personne est considérée en surpoids si son IMC (Indice de masse corporelle) est supérieur à 25. L'IMC se calcule de la manière suivante : masse en kg/(taille en m)^2.
La proportion de personnes en surpoids dans l'échantillon étudié est de 29,5%. Déterminer un intervalle de confiance au niveau de confiance de 95%.

3° Critiques de la notion de représentativité
a) Si l'échantillon n'avait été composé que d'une seule personne de 70 ans de sexe masculin, aurait-il été représentatif des variables genre et âge selon la règle de décision de la question 1° e).
b) Selon la même règle, un échantillon de 460 personnes peut-il être représentatif de tous les critères imaginables présents dans une population ?
c) En quoi l'intervalle de confiance calculé dans la seconde question aurait été modifié si l'échantillon n'avait pas été représentatif de l'âge et du sexe la population ?

Partie III : Problèmes et Compléments.

III. 1. Problèmes récapitulatifs

Exercice 46 : Loi exponentielle et propriété de mémoire (98 p.361, repères TS, Hachette)
On cherche dans le problème suivant à décrire la loi de la variable aléatoire X de densité *f* sur $[0;+\infty[$, représentant la durée de vie d'un noyau radioactif. Les données expérimentales ont mis en évidence les résultats suivants :

- Tous les noyaux ont une durée qui suivent la même loi ;
- La mort d'un noyau est indépendante de celles des autres noyaux ;
- Pour tous *t* et *h* dans , $P_{(X\geq t)}(t\leq X\leq t+h)=P(X\leq h)$.

On considère la fonction *F* définie sur $[0;+\infty[$ par : $F(t)=P(X\leq t)$.

Etude théorique.

1. Interpréter en terme de durée de vie, le résultat $P_{(X\geq t)}(t\leq X\leq t+h)=P(X\leq h)$.
2. Démontrer que pour tous *a* et *b* dans $[0;+\infty[$, on a : $P(a\leq X\leq b)=F(b)-F(a)$ et $P(X\geq a)=1-F(a)$.
3. En déduire que pour tous *t* et *h* dans $[0;+\infty[$, $F(t+h)-F(t)=F(h)\times(1-F(t))$.

On considère maintenant la fonction *G* définie sur $[0;+\infty[$ par : $G(t)=1-F(t)$.

4. Montrer que *G* est dérivable sur $[0;+\infty[$ et calculer sa dérivée.
5. Démontrer que pour tous *t* et *h* dans $[0;+\infty[$, $G(t+h)=G(t)\times G(h)$.
6. En déduire qu'il existe un réel λ tel que, pour tout *t* dans $[0;+\infty[$: $F(t)=1-e^{-\lambda t}$.
7. Montrer que $\lambda>0$ et en déduire que X suit une loi exponentielle de paramètre λ .

Application
On considère un noyau radioactif de carbone 14. Sa durée de vie X est une variable aléatoire qui suit une loi exponentielle de paramètre $\lambda=1{,}21\times10^{-4}$, où λ est appelée la constante radioactive par unité de temps, ici en années.

1. Calculer $P(X\geq1000)$ et interpréter ce résultat.
2. Déterminer le réel *t* tel que $P(X\geq t)=P(X\leq t)$. Ce réel *t* est appelé période radioactive ou demi-vie du noyau radioactif.

Exercice 47 : Conditionnement et loi exponentielle (99 p.361, repères TS, Hachette)

Un fabriquant de jouets vend un modèle de poupées qui« parle et marche » grâce à un mécanisme électronique. On appelle « durée de vie » d'une poupée le temps pendant lequel le mécanisme fonctionne correctement avant la première défaillance.
La variable aléatoire T, représentant la durée de vie exprimée en années d'une poupée prise au hasard dans la production, suit une loi exponentielle de paramètre $\frac{1}{3}$.
La probabilité $P(T\leq t)$ que la durée de vie de la poupée soit inférieure à *t* années est alors donnée par $P(T\leq t)=1-e^{-\frac{1}{3}t}$.
Dans cet exercice, pour chaque probabilité demandée, on donnera sa valeur exacte, puis une valeur approchée à 10^{-4} près.

1. a. Déterminer la probabilité *p* qu'une poupée ne fonctionne plus au bout d'une année.
b. Exprimer, en fonction de *t*, la probabilité $P(T>t)$ qu'une poupée n'ait aucune défaillance pendant *t* années.
2. J'ai acheté une poupée. On note A l'événement « la poupée n'a aucune défaillance pendant une année » et B l'événement « la poupée n'a aucune défaillance pendant trois ans ».
a. Déterminer les probabilités $P(A)$ et $P(B)$ des événements A et B.
b. Sachant que la poupée fonctionne parfaitement au bout d'un an, quelle est la probabilité $P_A(B)$ que la poupée fonctionne encore au bout de trois ans ? Justifier le calcul.

Exercices 48-51 : Conditionnement, loi exponentielle, binomiale, Moivre-Laplace
(125-126-126-128 p.412-413, repères TS, Hachette)

125. Une entreprise de confection possède 1 400 machines à coudre. On sait que la durée de vie (en années) de chaque machine à coudre suit la loi exponentielle de paramètre $\lambda = 0,2$.

1. Quelle est la durée de vie moyenne d'une machine à coudre ?

2. Quelle est la probabilité qu'une machine à coudre ait une durée de vie inférieure à 2 ans ?

Depuis sa création, l'entreprise est en difficulté. Au bout de deux ans, elle dépose le bilan et met en vente un lot composé de toutes ses machines à coudre. On suppose qu'elle n'a jamais remplacé les machines devenues hors-service pendant ses deux années d'activité. On note X le nombre de machines hors-service au moment de la vente, cette quantité étant inconnue de l'acheteur.

3. Quelle est la loi de X ? Donner son espérance et son écart-type.

4. À quel intervalle appartient approximativement la proportion de machines à coudre hors-service au moment de la vente, avec une probabilité de 95 % ?

5. En utilisant l'approximation fournie par le théorème de Moivre-Laplace, calculer la probabilité qu'un acheteur se retrouve avec plus de la moitié des machines à coudre qu'il a achetées hors-service.

126.

Partie A

Une entreprise effectue des contrôles pour détecter si un produit satisfait aux normes prévues. Le produit est conditionné en boîtes. Les contrôles montrent que la probabilité qu'une boîte prélevée au hasard dans la production soit défectueuse est égale à 0,006.
Soit n un entier naturel. On note X la variable aléatoire qui à chaque lot de n boîtes du produit tirées au hasard et avec remise dans la production associe le nombre de boîtes défectueuses dans ce lot.

1. Justifier que la variable aléatoire X suit une loi binomiale dont on précisera les paramètres. Donner l'espérance mathématique de X en fonction de n.

2. Déterminer, en fonction de n, la probabilité qu'il n'y ait aucune boîte défectueuse dans le lot.

3. Dans cette question, on prend $n = 10\,000$.
On admet que l'on peut approcher la loi de $\dfrac{X - 60}{7,72}$ par la loi normale centrée réduite. En déduire la probabilité qu'il y ait entre 50 et 70 boîtes défectueuses dans le lot de 10 000 boîtes.

Partie B

L'entreprise désire estimer la proportion inconnue p de clients satisfaits par son produit. Elle organise une enquête de satisfaction auprès d'un échantillon de 100 clients. Cette enquête a révélé que 85 d'entre eux étaient satisfaits. On considère que la taille de l'échantillon par rapport au nombre total des clients est suffisamment faible pour considérer que les personnes ont été tirées avec remise.

1. Proposer une estimation naturelle de la proportion inconnue p basée sur les résultats de l'enquête.

2. Donner une estimation de p par intervalle de confiance au niveau de confiance asymptotique 95 %. Arrondir les bornes à 10^{-2}.

3. Peut-on affirmer que p est compris dans cet intervalle de confiance ? Pourquoi ?

D'après BTS Opticien Lunetier, 2007.

127. Une usine fabrique, en grande quantité, des rondelles d'acier pour la construction. Leur diamètre est exprimé en millimètres.

Partie A

Une rondelle est conforme pour le diamètre lorsque celui-ci appartient à l'intervalle $[89,6 ; 90,4]$.

1. On note X la variable aléatoire qui, à chaque rondelle prélevée au hasard dans la production, associe son diamètre. On suppose que X suit la loi normale d'espérance 90 et d'écart-type $\sigma = 0,17$. Calculer la probabilité qu'une rondelle prélevée au hasard dans la production soit conforme.

2. L'entreprise désire améliorer la qualité de la production des rondelles : il est envisagé de modifier le réglage des machines produisant les rondelles. On note D la variable aléatoire qui, à chaque rondelle prélevée dans la production future, associera son diamètre. On suppose que D suit la loi normale d'espérance 90 et d'écart-type σ_1. Déterminer σ_1 pour que la probabilité qu'une rondelle prélevée au hasard dans la production future ait un diamètre conforme soit égale à 0,99.

Partie B

On note E l'événement : « une rondelle prélevée au hasard dans un stock important a un diamètre non conforme. » On suppose $P(E) = 0,02$. On prélève au hasard quatre rondelles dans le stock pour vérification de leur diamètre. Le stock est assez important pour que l'on puisse assimiler ce prélèvement à un tirage avec remise de quatre rondelles. On considère la variable aléatoire Y qui à tout prélèvement de quatre rondelles associe le nombre de rondelles de ce prélèvement ayant un diamètre non conforme.

1. Justifier que la variable Y suit une loi binomiale dont on déterminera les paramètres.
2. Calculer la probabilité que, dans un tel prélèvement, aucune rondelle n'ait un diamètre non conforme.
3. Calculer la probabilité que, dans un tel prélèvement, au plus une rondelle ait un diamètre non conforme.

Partie C

Les rondelles sont commercialisées par lots de 1 000. On prélève au hasard un lot de 1 000 dans un dépôt de l'usine. On assimile ce prélèvement à un tirage avec remise de 1 000 rondelles. On considère la variable aléatoire Z qui, à tout prélèvement de 1 000 rondelles, associe le nombre de rondelles non conformes parmi ces 1 000 rondelles.

1. Quelle est la loi suivie par Z ?
2. On admet que la loi de $\frac{Z-20}{4,43}$ peut être approchée par la loi normale centrée réduite. Justifier cette démarche.
3. À l'aide de cette approximation, calculer la probabilité qu'il y ait au plus 15 rondelles non conformes dans le lot de 1 000 rondelles.

D'après BTS Aéronautique, 2005.

128.

Partie A

Une entreprise A fabrique des composants électroniques. Durant leur montage, ces composants sont soumis à des températures de soudure d'au moins 200 degrés Celsius. On suppose que cette température est variable et vaut $T_A = 200 + E_A$ où E_A est une variable aléatoire suivant la loi exponentielle de paramètre $\lambda_A = 0,02$.

1. Que vaut l'espérance de E_A ?
2. Si la température de soudure est supérieure à une température critique de 340 degrés Celsius, le composant électronique est endommagé. Quelle est la probabilité qu'un composant électronique soit endommagé après soudure ?
3. Une autre entreprise B fabrique les mêmes composants électroniques, mais utilise une technique de soudure différente, pour laquelle la température est une variable aléatoire $T_B = 200 + E_B$, où E_B suit la loi exponentielle de paramètre $\lambda_B = 0,025$. Quelle est la probabilité qu'un composant électronique soit endommagé après soudure dans l'entreprise B ?

Partie B

Un fabricant de voitures utilise des composants électroniques achetés aux entreprises A et B pour assembler ses véhicules. Ce fabricant achète plus précisément 70 % de ses composants électroniques à A et 30 % à B. On admet dans cette partie que la probabilité p_A que les composants de l'entreprise A soient endommagés vaut 0,06, tandis que pour l'entreprise B, la probabilité p_B vaut 0,03.

1. On prélève au hasard un composant électronique utilisé par le fabricant de voitures. Montrer que la probabilité que le composant soit endommagé vaut 0,051.

Aide
On pourra représenter la situation par un arbre pondéré.

2. On prélève au hasard un lot de 500 composants électroniques. On note X le nombre de composants endommagés dans ce lot. Quelle loi suit la variable aléatoire X ?
3. Pourquoi est-il raisonnable d'affirmer que $\frac{X-25,5}{4,9}$ suit approximativement la loi normale centrée réduite ?
4. En utilisant cette approximation, calculer la probabilité que le lot contienne plus de 30 composants électroniques endommagés.

Exercice 52 : De la loi géométrique vers la loi exponentielle (84 p.428, math'x TS, Didier)

84 Temps d'attente de ROUGE TICE

Une urne contient 1 boule rouge et 9 boules blanches, indiscernables au toucher. Une partie consiste à répéter le tirage au hasard d'une boule de l'urne, avec remise après chaque tirage. La partie ne s'arrête que lorsque la boule rouge est tirée.
On désigne par N la variable aléatoire associant à une partie le nombre de tirages nécessaires à l'obtention de la boule rouge.

A. Modélisation discrète : loi géométrique

1. a. Quelles sont les valeurs prises par N ?
b. Calculer $P(N = 1)$, $P(N = 2)$, puis $P(N = k)$, pour $k \in \mathbb{N}^*$.
c. Calculer $P(N \leq n)$ et $P(N > n)$ pour $n \in \mathbb{N}^*$.
Calculer et interpréter $\lim\limits_{n \to +\infty} P(N > n)$.

2. Sur Geogebra, construire un histogramme dont les rectangles R_k, pour k entier naturel non nul, ont pour base $[k-1\,;k]$ et pour hauteur $p_k = P(N = k)$.
Pour la création de l'Histogramme, on pourra se limiter à $1 \leq k \leq 150$ et utiliser :
Histogramme [Hbornes, Hhauteurs] avec les listes :
Hbornes : Sequence[k, k, 0, 150] et
Hhauteurs : Sequence[« pk », k, 0, ~~150~~]. 149

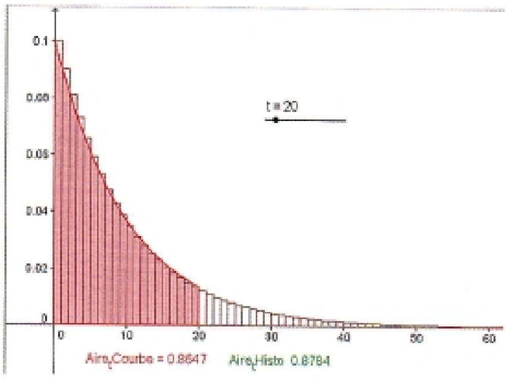
3. a. Que vaut l'aire d'un rectangle R_k ?
b. Quelles aires correspondent aux probabilités $P(N \leq 5)$, $P(10 < N \leq 20)$ et $P(N > 25)$?
c. En déterminer des valeurs approchées à 10^{-4} près.

B. Modélisation continue : loi exponentielle

L'histogramme construit évoque la courbe de densité d'une loi exponentielle.

1. Rappeler l'expression d'une telle fonction et déterminer son paramètre λ tel que $f(0) = 0{,}1$.
On prend dorénavant $\lambda = 0{,}1$.

2. a. Tracer sur le graphique la courbe de cette fonction de densité. Qu'observe-ton ?



b. Y faire figurer :
• un curseur où l'entier t varie de 1 à 150
• $\text{Aire}_t\text{Courbe}$: Integrale de f, de 0 à t
• $\text{Aire}_t\text{Histo}$: Somme [Hhauteurs, t]

c. Faire varier t avec le curseur et observer les valeurs de « $\text{Aire}_t\text{Courbe}$ » et « $\text{Aire}_t\text{Histo}$ ». Commenter.

3. On adopte ce modèle continu exponentiel pour approcher la loi de N.
a. Calculer $I(n) = \int_0^n f(x)\,dx$ et comparer avec la valeur de $P(N \leq n)$ obtenue à la question A.1.c.
b. En déduire des approximations des probabilités $P(N \leq 5)$, $P(10 < N \leq 20)$ et $P(N > 25)$, dont on donnera des valeurs approchées à 10^{-4} près.
c. Comparer ces approximations avec celles de A.3.c.

Exercice 53 : Intervalle de confiance issu de Moivre-Laplace (56 p.453, math'x TS, didier)

56 Intervalle de confiance « standard »

On considère une population où la proportion p d'un caractère est inconnue. F_n est la variable aléatoire associant à tout échantillon de taille n, prélevé au hasard et avec remise dans la population, la fréquence du caractère observée sur l'échantillon.
Selon le cours de Terminale, on obtient un intervalle de confiance de p au niveau de confiance de 95 % par une réalisation de l'intervalle $\left[F_n - \frac{1}{\sqrt{n}}\,;F_n + \frac{1}{\sqrt{n}}\right]$ dont les bornes sont aléatoires.
On montre dans le supérieur, qu'un intervalle de confiance « standard » à 95 % est fourni par une réalisation de l'intervalle :

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

1. Montrer que, pour toute réalisation f de la variable aléatoire F_n, l'intervalle de confiance « standard » à 95 % est inclus dans l'intervalle de confiance à 95 % au programme de Terminale.

2. Que peut-on en déduire quant à la probabilité de recouvrement de la valeur p par l'intervalle de confiance « standard » à 95 % ?

3. À l'aide d'un tableur, comparer les fréquences de recouvrement de p par les deux intervalles de confiance sur des échantillons de taille 100.
Pour exemple, sur l'image d'écran suivante, on a simulé 100 échantillons de taille 100 pour une valeur de p entrée en cellule B1.

B106 | =SI(ET(B1>=B104;B1<=B105);1;0)

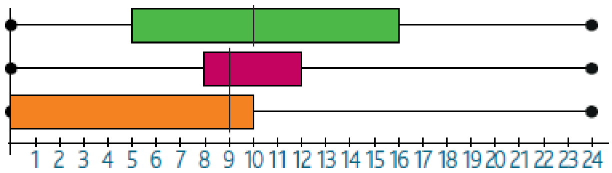
	A	B	C	D
99	97	1	1	0
100	98	1	0	0
101	99	0	0	0
102	100	1	1	0
103	**fréquence échantillon**	**0,74**	**0,63**	**0,69**
104	borne inf IC standard	0,65	0,54	0,6
105	borne sup IC standard	0,83	0,72	0,78
106	Recouvrement ?	1	1	1
107	fréquence de recouvrement IC standard	**96**	**%**	
108	borne inf IC terminale	0,64	0,53	0,59
109	borne sup IC terminale	0,84	0,73	0,79
110	Recouvrement ?	1	1	1
111	fréquence de recouvrement IC terminale	**99**	**%**	
112				

III. 2. Compléments

a) Statistiques descriptives

Exercice 54 (58 p.167 , repère 1ère S, Hachette)

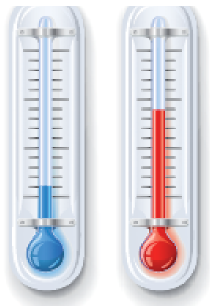
58. On étudie le nombre d'heures de pluie journalier de trois villes A, B et C pendant une année.
Voici les diagrammes en boîte résumant ces données
(De haut en bas : A, B puis C).



1. Dans quelle ville faut-il habiter si l'on souhaite qu'il pleuve le plus rarement possible ?
2. Dans quelle ville faut-il habiter si l'on souhaite qu'il pleuve le plus souvent possible ?

Exercice 55 (73 p.171 , repère 1ère S, Hachette)

73. Ni chaud ni froid



On étudie l'évolution de la température « moyenne » en France métropolitaine depuis 1900. Pour une année donnée, la température «moyenne » est la moyenne des températures maximales et minimales, relevées dans 22 stations du territoire et sur toute l'année. La température « normale » est une moyenne calculée sur une longue période de référence. Les températures sont données par leur écart par rapport à la température « normale ».

Exemple
- un écart égal à – 0,5 °C signifie que la température «moyenne » relevée est inférieure de 0,5 °C à la température normale.
- un écart égal à 0,8 °C signifie que la température «moyenne » relevée est supérieure de 0,8 °C à la température « normale ».

Ce tableau recueille les données de la période 1987-2007

Année	Écart à la normale en °C	Année	Écart à la normale en °C	Année	Écart à la normale en °C
1987	– 0,5	1994	1,2	2001	0,6
1988	0,3	1995	0,7	2002	1,0
1989	0,8	1996	– 0,3	2003	1,3
1990	0,9	1997	1,0	2004	0,5
1991	– 0,1	1998	0,4	2005	0,5
1992	0,2	1999	0,9	2006	1,1
1993	– 0,1	2000	1,0	207	0,8

1. a. Déterminer la moyenne (arrondie à 0,1 °C) de la série statistique des écarts à la normale.
b. Sachant que la température normale est 11,7 °C, quelle a été la température moyenne (arrondie à 0,1 °C) en France métropolitaine au cours de cette période ?

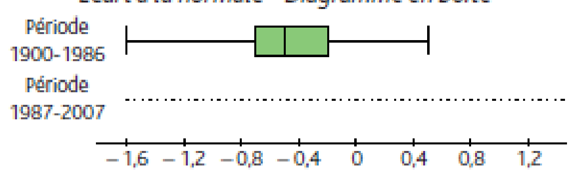
2. Recopier puis compléter le tableau ci-dessous :

Période 1987-2007 - Tableau d'effectifs

Écart à la normale en °C	–0,5	–0,3	–0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2	1,3
Nbre d'années (période 1987-2007)															
Eff. cumulés croissants															

3. Déterminer, en justifiant, la médiane et les quartiles de la série statistique des écarts à la normale.
Recopier, puis tracer le diagramme en boîtes de la série statistique dans le repère ci-dessous :

Écart à la normale - Diagramme en boîte



Justifier les affirmations suivantes :
a. Depuis 1987, plus de 50 % des années ont été plus chaudes que l'année la plus chaude de la période 1900-86.
b. Au moins 50 % des années de la période 1900-1986 ont été plus froides ou aussi froides que l'année la plus froide de la période 1987-2007.

D'après Bac

b) Lois discrètes

Exercice 56 : Activités autour de lancers de dés

Le jeu de données « Lancers_dés.ods » contient le résumé de 30 lancers de dé effectués par chaque élève d'une classe.

- 1ère activité possible : en supposant que tous les dés sont équilibrés, observer les fluctuations d'apparition de chaque face.
- 2ème activité possible : tester l'équilibre de chaque dé (plusieurs approches sont envisageables).

Exercices 57-58 (p.211, repères 1ère S, Hachette)

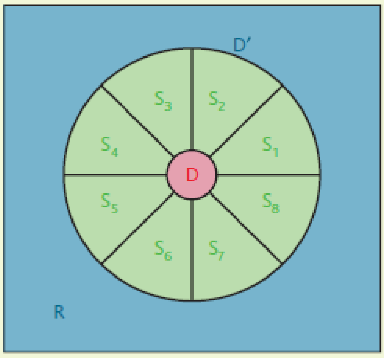
95. Un carré de côté 20 cm de centre O est partagé selon les zones suivantes :

- un disque D de centre O et de rayon 1 cm ;
- 8 secteurs $S_1, S_2, \ldots, S_8$ de même aire, délimités par les frontières du disque D et du disque D′ de même centre et de rayon 9 cm.
- une zone R entre le disque D′ et le bord du carré.

On place un point aléatoirement dans le carré. La probabilité de placer le point dans une zone quelconque du carré est proportionnelle à l'aire de cette zone.

1. a. Déterminer la probabilité $p(\text{D})$ pour que le point soit placé dans le disque D.

b. Déterminer la probabilité $p(\text{S}_1)$ pour que le point soit placé dans le secteur S_1.



2. Pour cette question on utilisera les valeurs approchées suivantes : $p(\text{D}) = 0{,}008$ et pour tout $k \in \{1, 2, 3, 4, 5, 6, 7, 8\}$, $p(\text{S}_k) = 0{,}0785$.

À cette situation aléatoire est associé le jeu suivant :

- un point placé dans le disque D fait gagner 10 €.
- un point placé dans le secteur S_k fait gagner k € pour tout k appartenant à $\{1, 2, 3, 4, 5, 6, 7, 8\}$;
- un point placé dans la zone R fait perdre 4 €.

On note X la variable aléatoire égale au gain algébrique obtenu.

Calculer la probabilité $p(\text{R})$ pour que le point soit placé dans la zone R. Calculer l'espérance de X.

D'après de Bac S.

96. Un sac contient trois boules numérotées respectivement 0,1 et 2, indiscernables au toucher. On tire une boule du sac, on note son numéro x et on la remet dans le sac ; puis on tire une seconde boule, on note son numéro y et on la remet dans le sac. Toutes les boules ont la même probabilité d'être tirées. À chaque tirage de deux boules, on associe dans le plan, muni d'un repère orthonormal $(\text{O} ; \vec{\imath}, \vec{\jmath})$, le point M de coordonnées $(x ; y)$. On désigne par D le disque de centre O et de rayon 1,7. Les résultats seront donnés sous forme de fraction irréductible.

1. Placer dans le plan muni du repère $(\text{O} ; \vec{\imath}, \vec{\jmath})$ les points correspondant aux différents résultats possibles.

2. Calculer la probabilité de chacun des événements suivants :

A : « Le point M est sur l'axe des abscisses » ;

B : « Le point M appartient au cercle de centre O et de rayon 1 ».

3. Soit X la variable aléatoire qui, à chaque tirage de deux boules, associe la somme $x^2 + y^2$.

a. Déterminer la loi de probabilité de la variable aléatoire X. Calculer son espérance mathématique E(X).

b. Montrer que la probabilité de l'événement « le point M appartient au disque D » est égale à $\frac{4}{9}$.

D'après Amérique du sud.

Exercice 59-60 (94 p.359, repères TS, Hachette ; 103 p.256, 1ère S, Hachette)

94. BAC Pour réaliser une enquête, un employé interroge des personnes prises au hasard dans une galerie commerçante. Il se demande si trois personnes au moins accepteront de répondre.

1. Dans cette question, on suppose que la probabilité qu'une personne choisie au hasard accepte de répondre est égale à 0,1. L'employé interroge 50 personnes de manière indépendante. On considère les événements A : « Au moins une personne accepte de répondre », B : « Strictement moins de trois personnes acceptent de répondre » et C : « Trois personnes acceptent de répondre ».
Calculer les probabilités des événements A, B et C.
On arrondira au millième.

2. Soit n un entier naturel supérieur ou égal à 3. Dans cette question, on suppose que la variable aléatoire X qui, à tout groupe de n personnes interrogées indépendamment, associe le nombre de personnes ayant accepté de répondre, suit la loi définie pour tout entier k tel que $0 \leqslant k \leqslant n-1$, par :

$$P(X=k)=\frac{e^{-a}a^k}{k!} \quad \text{et} \quad P(X=n)=1-\sum_{k=1}^{n-1}\frac{e^{a}a^k}{k!}$$

$\left(\text{formules dans lesquelles } a=\frac{n}{10}\right)$.

a. Montrer que la probabilité qu'au moins trois personnes répondent est donnée par $f(a)=1-e^{-a}\left(1+a+\frac{a^2}{2}\right)$.

b. Calculer $f(5)$. En donner l'arrondi au millième. Cette modélisation donne-t-elle un résultat voisin de celui obtenu à la question 1. ?

3. On conserve le modèle de la question 2.. On souhaite déterminer le nombre minimum de personnes à interroger pour que la probabilité que trois d'entre elles au moins répondent soit supérieure ou égale à 0,95.

a. Étudier les variations de la fonction f définie sur $\mathbb{R}^*$ par $f(x)=1-e^{-x}\left(1+x+\frac{x^2}{2}\right)$ ainsi que sa limite en $+\infty$. Dresser son tableau de variations.

b. Montrer que l'équation $f(x)=0{,}95$ admet une unique solution sur $\mathbb{R}^+$, et que cette solution est comprise entre 6,29 et 6,3.

c. En déduire le nombre minimum de personnes à interroger.

103. Les sangs mélangés

Un laboratoire pharmaceutique a mis au point un test sanguin de dépistage d'une maladie.
On sait que la probabilité qu'un individu soit atteint de cette maladie est p.
On considère un groupe de n individus testés pour cette maladie. Soit X la variable aléatoire correspondant au nombre de tests positifs.

1. Quelle loi suit X ? Donner ses paramètres.

2. Quelle est la valeur de $p(X \geqslant 1)$?

3. Afin de minimiser le coût des tests, le laboratoire fait un premier test en regroupant les n échantillons de sangs prélevés.
Si ce test est négatif, alors le laboratoire peut être sûr que les n individus testés ne sont pas atteints par la maladie. Le laboratoire a eu besoin d'un seul test pour connaître le résultat des n individus.
Dans le cas contraire, le laboratoire réalise un test sur chaque échantillon prélevé. Le laboratoire a alors réalisé en tout $n+1$ tests.
Soit Y la variable aléatoire correspondant au nombre de tests réalisés par le laboratoire.
a. Donner les valeurs que peut prendre Y.
b. Quelle loi suit Y ?
c. Démontrer que $E(Y)=n+1-n(1-p)^n$.

4. Soit Z la variable aléatoire correspondant au nombre de tests réalisés sur un seul échantillon de sang.
On admet $Z=\frac{Y}{n}$.
a. Donner les valeurs que peut prendre Z.
b. Quelle loi suit Z ?
c. Démontrer que $E(Z)=1-(1-p)^n+\frac{1}{n}$.

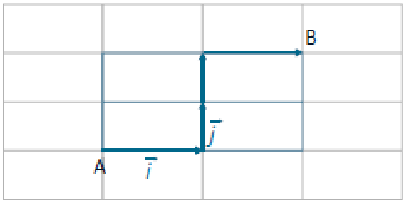
5. On admet que $p=0{,}009$.
Soit (u_n) la suite définie sur $\mathbb{N}^*$ par :
$$u_n=1-(1-p)^n+\frac{1}{n}.$$
a. Représenter graphiquement les premiers termes de cette suite à l'aide de la calculatrice.
b. En déduire la valeur de n pour laquelle u_n admet la plus petite valeur.
c. En déduire la valeur de n pour laquelle le coût par individu est le moins important. Chaque test coute 10 € au laboratoire. En pratiquant cette méthode, quelle est l'économie réalisée par le laboratoire (en euros par individus) ?

Exercice 61 : Marche aléatoire avec sauts indépendants (104 p.257, 1ère S, Hachette)

104. Soit $(A ; \vec{i}, \vec{j})$ un repère du plan et B(2 ; 2).

Partie 1

On considère un trajet aléatoire de longueur 4 au départ du point A. Chaque trajet est constitué d'une succession de quatre vecteurs ($\vec{i}$ ou $\vec{j}$).
Par exemple, on a représenté ci-dessous le trajet $(\vec{i}, \vec{j}, \vec{j}, \vec{i})$, qui va de A à B.



On modélise un trajet par une succession de quatre expériences identiques et indépendantes. Chaque expérience a deux issues possibles : vecteur $\vec{i}$ ou vecteur $\vec{j}$.
Soit X la variable aléatoire correspondant au nombre de vecteurs $\vec{i}$ contenus dans un trajet de longueur 4.

1. Quelle loi suit X ? Donner ses paramètres.

2. Quelle est la probabilité d'arriver au point B à la fin du trajet ?

Partie 2

Soit C(4 ; 5). On étudie les trajets aléatoires partant du point A et de longueur 9.
Soit Y la variable aléatoire correspondant au nombre de vecteurs $\vec{i}$ contenus dans un trajet de longueur 9.

1. Quelle loi suit Y ? Donner ses paramètres.

2. Quelle est la probabilité d'arriver au point C à la fin du trajet ?

3. Quelle est la probabilité d'arriver au point C et de passer par le point B ?

c) Conditionnement

Exercice 62 : Conditionnement et marche aléatoire (97 p.360 , repères TS, Hachette)

97. BAC On étudie le mouvement aléatoire d'une puce. Cette puce se déplace sur trois cases notées A, B et C. À l'instant 0, la puce est en A.

Pour tout entier naturel n :

- si à l'instant n la puce est en A, alors à l'instant $(n+1)$, elle est :
  - soit en B avec une probabilité égale à $\frac{1}{3}$;
  - soit en C avec une probabilité égale à $\frac{2}{3}$;
- si à l'instant n la puce est en B, alors à l'instant $(n+1)$, elle est :
  - soit en C ;
  - soit en A de façon équiprobable ;
- si à l'instant n la puce est en C, alors elle y reste.

On note A_n (respectivement B_n et C_n) l'événement « À l'instant n, la puce est en A » (respectivement en B et en C).
On note a_n (respectivement b_n et c_n) la probabilité de l'événement A_n (respectivement B_n et C_n).
On a donc $a_0 = 1$, $b_0 = c_0 = 0$.
Pour traiter l'exercice, on pourra s'aider d'un arbre pondéré.

1. Calculer a_k, b_k et c_k, pour k entier naturel tel que :
$$1 \leqslant k \leqslant 3.$$

2. a. Montrer que, pour tout entier naturel n :
$$a_n + b_n + c_n = 1 \quad \text{et} \quad \begin{cases} a_{n+1} = \frac{1}{2} b_n \\ b_{n+1} = \frac{1}{3} a_n \end{cases}.$$

b. Montrer que pour tout entier naturel n, $a_{n+2} = \frac{1}{6} a_n$.

c. En déduire que pour tout entier naturel p :
$$\begin{cases} a_{2p} = \left(\frac{1}{6}\right)^p \\ b_{2p} = 0 \end{cases} \quad \text{et} \quad \begin{cases} a_{2p+1} = 0 \\ b_{2p+1} = \frac{1}{3}\left(\frac{1}{6}\right)^p \end{cases}.$$

3. Montrer que $\lim\limits_{n \to +\infty} a_n = 0$.

On admet que $\lim\limits_{n \to +\infty} b_n = 0$. Quelle est la limite de c_n lorsque n tend vers $+\infty$?

D'après Bac, 2005

Exercice 63 : Sensibilité et spécificité d'un test de diagnostic (p.344 , repères TS, Hachette)

Exercice résolu 2 TICE Pertinence d'un test de dépistage

énoncé Dans une population donnée suffisamment grande, la proportion d'individus atteints d'une certaine maladie est x. Un laboratoire pharmaceutique informe sur les caractéristiques de son test spécifique à cette maladie :

- la probabilité qu'un individu malade ait un test positif est 0,99 ;
- la probabilité qu'un individu sain ait un test positif est 0,01.

On choisit un individu au hasard dans cette population et on le soumet au test. On note :

- M l'événement : « L'individu est malade » ;
- T l'événement : « Le test est positif » ;
- $f(x)$ la probabilité qu'une personne ayant un test positif soit malade.

1. Sur quel intervalle I est définie la fonction f ? Montrer que $f(x) = \frac{99x}{98x + 1}$, puis étudier son sens de variation sur I. Tracer $\mathcal{C}_f$ dans un repère orthonormé (prendre 10 cm comme unités en abscisses et 10 cm en ordonnées).

2. On considère que le test est fiable lorsque la probabilité qu'un individu ayant un test positif soit malade est supérieure ou égale à 0,95.
   a. Le test est-il fiable lorsque la proportion d'individus atteints par la maladie est de 5 % ?
   b. À l'aide d'un logiciel de calcul formel, déterminer la proportion d'individus atteints par la maladie à partir de laquelle le test est fiable.

3. a. Déterminer $g(x)$, où g est la fonction qui à x associe la probabilité qu'un individu dont le test est négatif ne soit pas malade.
   b. Pour pouvoir effectuer une campagne de dépistage sans inquiéter la population, on veut que cette probabilité soit supérieure à 0,99.
   À l'aide d'un logiciel de calcul formel, déterminer la proportion d'individus atteints par la maladie pour laquelle on n'inquiètera pas la population.

4. Conclure sur la qualité du test.

d) Techniques de simulation de variables aléatoires

Exercice 64 : Simulation d'une loi de Bernoulli (77 p.357, repères, TS, Hachette.)

77. On admet que l'on puisse assimiler la fonction « random » d'une calculatrice a une variable aléatoire suivant une loi uniforme sur $[0 ; 1]$.
Soit p un nombre réel dans $[0 ; 1]$.

1. Calculer $P(\text{random} < p)$.
2. On considère l'algorithme ci-dessous :

```
Variables
p un réel, X un entier
Début
Lire p
Si random < p
        Alors
        X → 1
        Sinon
        X → 0
FinSi
Afficher X
Fin
```

Quelle est la loi de la variable aléatoire X simulée par cet algorithme ?

3. En déduire un algorithme simulant une variable aléatoire suivant une loi de binomiale de paramètres n et p qui seront entrés par l'utilisateur.

Exercice 65 : Simulation d'une loi exponentielle (73 p.426, math'x TS, Didier)

73 Simulation d'une loi exponentielle

1. Soit U une variable aléatoire suivant la loi uniforme sur $[0\,;1[$. On considère la variable aléatoire $T = -\frac{1}{\lambda}\ln(1-U)$.

a. Justifier que T est bien définie et prend ses valeurs dans $[0\,;+\infty[$.

b. Montrer que, pour $t \geqslant 0$, $P(T \leqslant t) = P(U \leqslant 1 - e^{-\lambda t})$; en déduire l'expression de $P(T \leqslant t)$ en fonction de t.

c. On désigne par f la fonction de densité de la variable aléatoire T.

Que vaut, pour $t \geqslant 0$, $F(t) = \int_0^t f(x)\,dx$?

En déduire l'expression de $f(t)$ pour $t \geqslant 0$ et conclure

2. Que simule sur un tableur l'instruction :

$\boxed{\text{= – 0,125*ln(1–ALEA())}}$?

Pourquoi peut-on aussi utiliser, plus simplement :

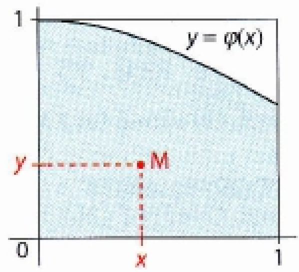
$\boxed{\text{= – 0,125*ln(ALEA())}}$?

3. Comment peut-on simuler, à l'aide d'une calculatrice ou d'un tableur, une série de réalisations d'une variable aléatoire de loi exponentielle de paramètre $\lambda = 0{,}005$?
Effectuer une simulation d'une telle série de 10 valeurs.

e) Méthode de Monte-Carlo

Exercice 66 (51 p.452, math'x TS, Didier)

51 Confiance en Monte Carlo ALGORITHMIQUE



Soit φ la fonction définie sur $[0\,;1]$ par $\varphi(x) = e^{-\frac{x^2}{2}}$ et soit $\mathscr{C} : y = \varphi(x)$.
On prend au hasard un point $M(x\,;y)$ où x et y sont des réels de $[0\,;1]$.

1. a. Montrer que la probabilité que M soit situé sous la courbe $\mathscr{C} : y = \varphi(x)$ est égale à $p = \int_0^1 \varphi(x)\,dx$.

b. Déterminer une valeur approchée de p à 10^{-5} près, à l'aide d'une calculatrice ou d'un logiciel.

2. L'algorithme de Monte-Carlo, programmé ci-dessous sur Scilab, simule le choix au hasard de n points M et fournit une valeur approchée f de p.
Décrire précisément la simulation effectuée et expliquer en particulier le test de la ligne 5.

```
1 s=0;
2 n=input("nombre de points n = ");
3 for i=1:n
4     x=rand();y=rand();
5     if y <=exp(-x^2/2) then s=s+1;
6     end
7 end
8 afficher(s/n)
```

3. a. Donner, à partir de la fréquence f obtenue par cet algorithme, l'expression d'un intervalle de confiance de p au niveau de confiance de 95 %.

b. Rappeler la signification d'un tel intervalle.

4. a. On exécute 4 fois l'algorithme pour $n = 10\,000$ et on obtient : 0,853 4 ; 0,856 ; 0,856 1 et 0,861 7.
Combien a-t-on de décimales exactes ?

b. Pour quelles valeurs de n obtient-on, au niveau de confiance de 95 %, une précision (donnée par la longueur de l'intervalle de confiance) égale à 10^{-3} ?

c. Adapter l'algorithme précédent et obtenir une valeur approchée de p à 10^{-3} près.

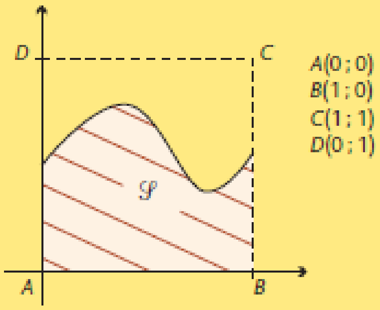
Exercice 67-69 (p.367, repères TS, Hachette)

méthode de Monte-Carlo

La méthode de Monte-Carlo est une méthode qui utilise les probabilités pour effectuer des calculs approchés de surfaces ou de volumes.

Pour appliquer cette méthode, nous utiliserons le protocole suivant :

- on tire un grand nombre de points au hasard dans la carré *ABCD* (*cf.* figure ci-dessous ; on peut imaginer qu'on lance aléatoirement des fléchettes dans le carré *ABCD*) ;
- on considère une fonction F1 définie sur $[0\,;\,1]$ et à valeurs dans $[0\,;\,1]$;
- on calcule la proportion p de points situés à l'intérieur du domaine $\mathcal{S}$ (*cf.* ci-dessous, la surface hachurée), défini par :
  $\mathcal{S} = \{(x\,;\,y) \in [0\,;\,1] \times [0\,;\,1] \,/\, y \leqslant F_1(x)\}$;
- on admet que p est une valeur approchée de $\mathcal{S}$.



122. Approximation de π

On considère le quart de cercle trigonométrique supérieur droit.

1. Déterminer $F_1(x)$.
2. Recopier et compléter le programme *Algobox* ci-après en indiquant dans l'onglet « utiliser une fonction numérique », la fonction trouvée au 1..
3. Expliquer le rôle des variables C, x et y.
4. Lancer ce programme pour les valeurs de N suivantes :
   1 000, 10 000, 100 000, 500 000.
5. De quel nombre ce programme calcule-t-il des valeurs approchées ? Comparer ces résultats avec celui de la calculatrice pour une validation expérimentale de la méthode.

```
VARIABLES
  N EST_DU_TYPE NOMBRE
  x EST_DU_TYPE NOMBRE
  y EST_DU_TYPE NOMBRE
  i EST_DU_TYPE NOMBRE
  aire EST_DU_TYPE NOMBRE
  C EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  LIRE N
  AFFICHER "Le nombre de lancer est égal à "
  AFFICHER N
  C PREND_LA_VALEUR 0
  POUR i ALLANT_DE 1 A N
    DEBUT_POUR
    x PREND_LA_VALEUR random()
    y PREND_LA_VALEUR random()
    SI (y<=F1(x)) ALORS
      DEBUT_SI
      C PREND_LA_VALEUR C+1
      TRACER_POINT (x,y)
      FIN_SI
    FIN_POUR
  aire PREND_LA_VALEUR C/N
  AFFICHER "Une valeur approchée de l'aire est "
  AFFICHER aire
FIN_ALGORITHME
```

123. Calcul de probabilité

On considère la variable aléatoire X de densité la fonction $F_1(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ définie sur $\mathbb{R}$. On admet que c'est une densité de probabilité (*cf.* chapitre 8), dont on ne connaît pas de primitives (sans utiliser le symbole $\int$).

1. Interpréter graphiquement le calcul de $P(0 \leqslant X \leqslant 1)$.
2. Adapter le programme de l'exercice 120., pour trouver une valeur approchée de $P(0 \leqslant X \leqslant 1)$.
3. Comparer cette valeur avec celle fournie par la calculatrice.

124. Calcul d'aire.

On considère la fonction suivante définie sur :

$$[0\,;\,1] \setminus \left\{\frac{1}{2}\right\} \text{ par } F_1(x) = \frac{1}{2}\left(\sin\left(\frac{1}{x-0{,}5}\right) + 1\right).$$

On souhaite déterminer l'aire du domaine $\mathcal{S}$.

1. Représenter cette fonction à l'aide d'une calculatrice ou d'un logiciel.
2. D'après le cours, peut-on considérer $\int_0^1 f(t)\,dt$ pour calculer l'aire sous la courbe ?
3. Adapter le programme de l'exercice 120., et lancer le pour des valeurs de N suivantes :
   10 000, 100 000 et 500 000.
4. Le résultat obtenu paraît-il satisfaisant ? Justifier.

Remarque
Il existe des généralisations de la notion d'intégrale permettant d'aborder ce type de calcul.