Loi normale
Échantillonnage et estimation

Christophe ROSSIGNOL*

Année scolaire 2019/2020

Table des matières

Table des figures

Liste des tableaux

1 Rappels sur la loi binomiale

1.1 Épreuve de Bernoulli

Définition : On appelle épreuve de BERNOULLI toute épreuve à deux issues possibles : un succès (noté S) ou un échec (noté $\overline{S}$).
La probabilité d'un succès $p = P(S)$ est appelé paramètre de l'épreuve de BERNOULLI.

Exemple : On lance un dé équilibré à six faces, les faces étant numérotés de 1 à 6.
On considère qu'il y a un succès lorsque le résultat du lancer est un 6, un échec sinon.
Il s'agit d'une épreuve de BERNOULLI de paramètre $\frac{1}{6}$.
On peut la représenter par l'arbre de la figure 1.

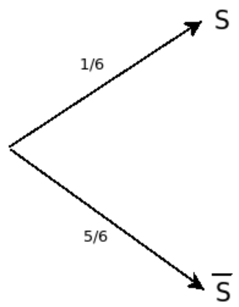


FIGURE 1: Un exemple d'épreuve de BERNOULLI

1.2 Schéma de Bernoulli – Loi binomiale

Définition :
- On appelle schéma de BERNOULLI l'expérience consistant à répéter n fois de manière indépendantes la même épreuve de Bernoulli de paramètre p.
- La loi binomiale de paramètres n et p est la loi de probabilité de la variable aléatoire X prenant prenant comme valeurs le nombre de succès (S) obtenus au cours des n épreuves du schéma de Bernoulli.
- On dit aussi que loi loi de probabilité de la variable aléatoire X suit la loi binomiale de paramètres n et p.

Exemple : On répète 2 fois de manière identiques et indépendantes l'épreuve de BERNOULLI de l'exemple précédent.
On note X la variable aléatoire qui donne le nombre de 6 obtenus. X suit la loi binomiale de paramètres 2 et $\frac{1}{6}$.
Le schéma de Bernoulli correspondant est donné sur la figure 2.

On a alors :
- $P(X = 0) = P\left(\bar{S}\bar{S}\right) = \frac{5}{6} \times \frac{5}{6} = \frac{25}{36}$
- $P(X = 1) = P\left(S\bar{S}\right) + \left(\bar{S}S\right) = \frac{1}{6} \times \frac{5}{6} + \frac{5}{6} \times \frac{1}{6} = \frac{5}{36} + \frac{5}{36} = \frac{10}{36}$
- $P(X = 2) = P(SS) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$

Remarques :
1. Si X suit la loi binomiale de paramètres n et p, X prend les valeurs 0, 1, 2, ..., n.
2. On peut toujours représenter un schéma de Bernoulli par un arbre pour calculer $P(X = k)$. Mais si n est grand, cela peut être fastidieux... On peut alors utiliser la calculatrice ou le tableur. Voir tableau 1.

Propriété : Soit X une variable aléatoire suivant la loi binomiale de paramètres n et p.
L'espérance de cette variable aléatoire est $E(X) = np$.

Exercices : Exercices 1, 2 de la feuille polycopiée « Exercices - loi binomiale ».

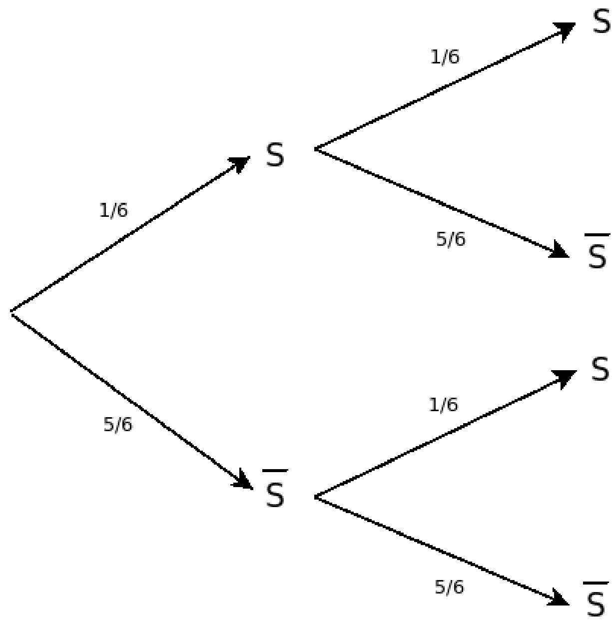


FIGURE 2: Un exemple de Schéma de BERNOULLI

	Casio	Texas	Tableur Open Office ou Excel
Syntaxe	Touche OPTN, puis choisir *STAT*, puis *DIST*, puis *BINM*, puis *Bpd* ou *Bcd* (voir p. 158)	Menu distrib (2nde var), puis choisir *binomFdp(* ou *binomFrép(* (voir p. 158).	Fonction LOI.BINOMIALE
$P(X=k)$	BinominalPD(k,n,p)	binomFdp(n,p,k)	=LOI.BINOMIALE(k;n;p;0)
$P(X \leqslant k)$	BinominalCD(k,n,p)	binomFRép(n,p,k)	=LOI.BINOMIALE(k;n;p;1)

TABLE 1: Utilisation de la calculatrice ou du tableur pour une loi binomiale

2 Loi normale

Activité : Exercice 3 de la feuille polycopiée « Exercices - loi binomiale ».

2.1 Courbe « en cloche »

Le diagramme en bâtons d'une loi binomiale de paramètres n et p, lorsque n est très grand et que p n'est proche ni de zéro ni de 1, peut être approché par une courbe « en cloche » (voir figure 3).
Cette courbe « en cloche » a les propriétés suivantes :

Propriétés : Courbe « en cloche »

— C'est la courbe représentative d'une fonction définie sur $\mathbb{R}$.
— L'aire totale comprise entre la courbe « en cloche » et l'axe des abscisses vaut 1.
— Elle dépend de deux paramètres nommés μ (mu) et σ (sigma). μ est appelé espérance et σ est appelé écart-type.
— Elle admet comme axe de symétrie la droite d'équation $x = \mu$ (voir figure 3).
— Plus σ est élevé, plus la courbe est « écrasée » autour de l'axe des abscisses (voir figure 3).

Remarque : On va définir, à l'aide de ces courbes « en cloche », une nouvelle loi de probabilité, pour des variables aléatoires *prenant toutes les valeurs réelles*.

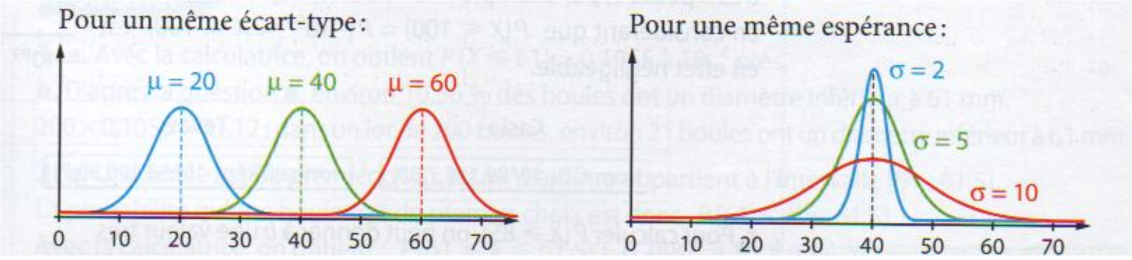


FIGURE 3: Courbes « en cloche »

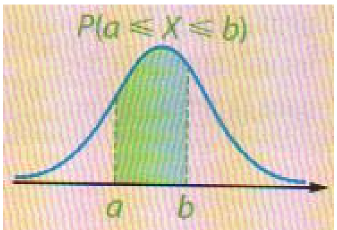
2.2 Loi normale

Définition : On considère une courbe « en cloche » de paramètres μ et σ et X une variable aléatoire prenant toutes les valeurs réelles.
On dit que X suit la loi normale de paramètres μ et σ si, pour tout nombres a, b, avec $a < b$:

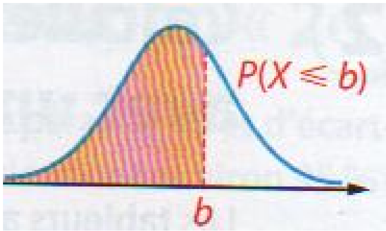
— La probabilité $P(a \leq X \leq b)$ que la variable aléatoire prenne des valeurs dans l'intervalle $[a\,;\,b]$ est égale à l'aire du domaine compris entre l'axe des abscisses, la courbe « en cloche » et les droites verticales d'équations $x = a$ et $x = b$. (voir figure 4a)

— La probabilité $P(X \leq b)$ que la variable aléatoire prenne des valeurs dans l'intervalle $[b\,;+\infty[$ est égale à l'aire du domaine compris entre l'axe des abscisses, la courbe « en cloche » et situé à gauche de la droite verticale d'équation $x = b$. (voir figure 4b)

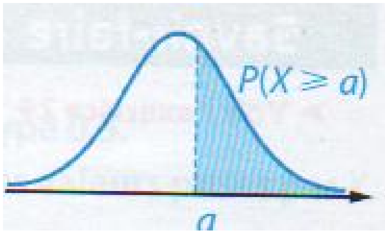
— La probabilité $P(X \geq a)$ que la variable aléatoire prenne des valeurs dans l'intervalle $]-\infty\,;\,a]$ est égale à l'aire du domaine compris entre l'axe des abscisses, la courbe « en cloche » et situé à droite de la droite verticale d'équation $x = a$. (voir figure 4c)



(a) $P(a \leq X \leq b)$



(b) $P(X \leq b)$



(c) $P(X \geq a)$

FIGURE 4: Loi normale de paramètres μ et σ

Remarques : 1. La fonction dont la courbe représentative est la courbe « en cloche » est alors appelé fonction de densité.

2. On a alors $P(X = a) = P(a \leq X \leq a) = 0$ donc $P(X > a) = P(X \geq a)$. Les inégalités peuvent être notées indifféremment larges ou strictes, cela ne change pas le probabilités.

3. Par un raisonnement graphique simple, on obtient les propriétés suivantes :

Propriété : Soit X une variable aléatoire qui suit la loi normale de paramètres μ et σ.

— $P(X \leq \mu) = P(X \geq \mu) = 0,5$

— $P(X \geq a) = 1 - P(X < a) = 1 - P(X \leq a)$

— $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = P(X \leq b) - P(X \leq a)$

Remarque : On utilisera la calculatrice pour déterminer des probabilités de variables aléatoires suivant une loi normale (voir tableau 2).

Exercices : 2, 3 page 138[1] – 8, 9 page 139[2] – 11 page 139 ; 12 page 140 et 28, 29 page 145[3] [Intervalle]

1. Utilisation de la courbe.
2. Utilisation de la calculatrice.
3. Applications.

Propriété : Intervalle de fluctuation (voir figure 5)
Soit X une variable aléatoire qui suit la loi normale de paramètres μ et σ.
On a alors :

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \simeq 0,95$$

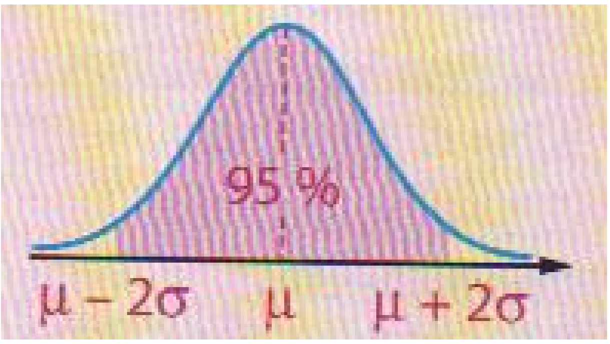


FIGURE 5: Intervalle de fluctuation

Remarque : Cela signifie qu'environ 95 % des valeurs prises par X sont dans l'intervalle $[\mu - 2\sigma\,;\, \mu + 2\sigma]$.
Exercices : 5, 6 page 139[4] [Intervalle]

2.3 Approximation d'une loi binomiale par une loi normale

Si n est très grand et que p n'est proche ni de zéro ni de 1, la loi binomiale de paramètres n et p peut être approchée par la loi normale de paramètres $\mu = np$ et $\sigma = \sqrt{np(1-p)}$.

3 Échantillonnage – Estimation

3.1 Définition – Utilisation

Rappel : On appelle échantillon de taille n la série statistique formée des résultats obtenus lorsqu'on répète n fois une expérience dans les mêmes conditions.
- Les distributions de fréquences varient d'un échantillon à l'autre pour la même expérience. C'est ce qu'on appelle la fluctuation d'échantillonnage.
- Même pour des échantillon de même taille, la distribution de fréquences peut varier.
- Lorsque la taille de l'échantillon augmente, les distributions de fréquences ont tendance à se stabiliser.

Remarque : Comme on répète dans les mêmes conditions une expérience n fois, on peut assimiler cet échantillon à une loi binomiale de paramètres $n$et p, où p est la proportion du caractère étudié dans la population totale. La distribution de fréquence de cet échantillon peut alors être assimilée à la loi de fréquence F_n.

Définition : Soit X_n une variable aléatoire qui suit le loi binomiale de paramètres n et p et $F_n = \frac{X_n}{n}$ la fréquence de succès.
On dit que l'intervalle I_n est un intervalle de fluctuation de F_n au seuil de 95 % si :

$$P(F_n \in I_n) \geq 0,95$$

Remarque : On utilise donc les intervalles de fluctuation dans les deux cas suivants :
- on connaît la proportion p de présence du caractère dans la population ;
- on fait une hypothèse sur la valeur de cette proportion et on veut valider (ou invalider) cette hypothèse (on parle alors de prise de décision).

4. Intervalle de fluctuation.

3.2 Échantillonnage – Prise de décision

Propriété : Soit un caractère dont la proportion dans une population donnée est p. On considère un échantillon de taille n.
Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, l'intervalle :

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

est un intervalle de fluctuation au seuil de 95 %.

On considère une population dans laquelle on suppose que la proportion d'un caractère est p.
On observe la fréquence f_{obs} de ce caractère dans un échantillon de taille n et on considère l'hypothèse « la proportion de ce caractère dans la population est p ».

On considère que les conditions $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ sont remplies et on note $I_n = \left[p - \frac{1}{\sqrt{n}} \,;\, p + \frac{1}{\sqrt{n}}\right]$ l'intervalle de fluctuation au seuil des 95 %.

On a alors la règle de décision suivante :
- Si $f_{\text{obs}} \in I_n$: on considère que l'hypothèse n'est pas remise en question et l'on accepte au seuil de risque de 5 % ;
- Si $f_{\text{obs}} \notin I_n$: on rejette l'hypothèse au seuil de risque de 5 % (ce qui signifie que le risque d'erreur par rejet à tort de l'hypothèse est d'environ 5 %).

Exercices : 13, 14 page 140 et 31, 32 page 145[5] [Intervalle]

3.3 Intervalle de confiance

On considère maintenant le cas où la proportion p du caractère dans la population totale est inconnue.
On veut estimer p à l'aide d'un échantillon de taille n, et on supposera que les conditions $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ sont remplies.

Définition : On observe une fréquence f_{obs} sur un échantillon de taille n.
On appelle intervalle de confiance de p au niveau de confiance de 95 % l'intervalle $\left[f_{\text{obs}} - \frac{1}{\sqrt{n}} \,;\, f_{\text{obs}} + \frac{1}{\sqrt{n}}\right]$.

Remarques : Cela signifie que la proportion inconnue a plus de 95 % de chances de se trouver dans cet intervalle.

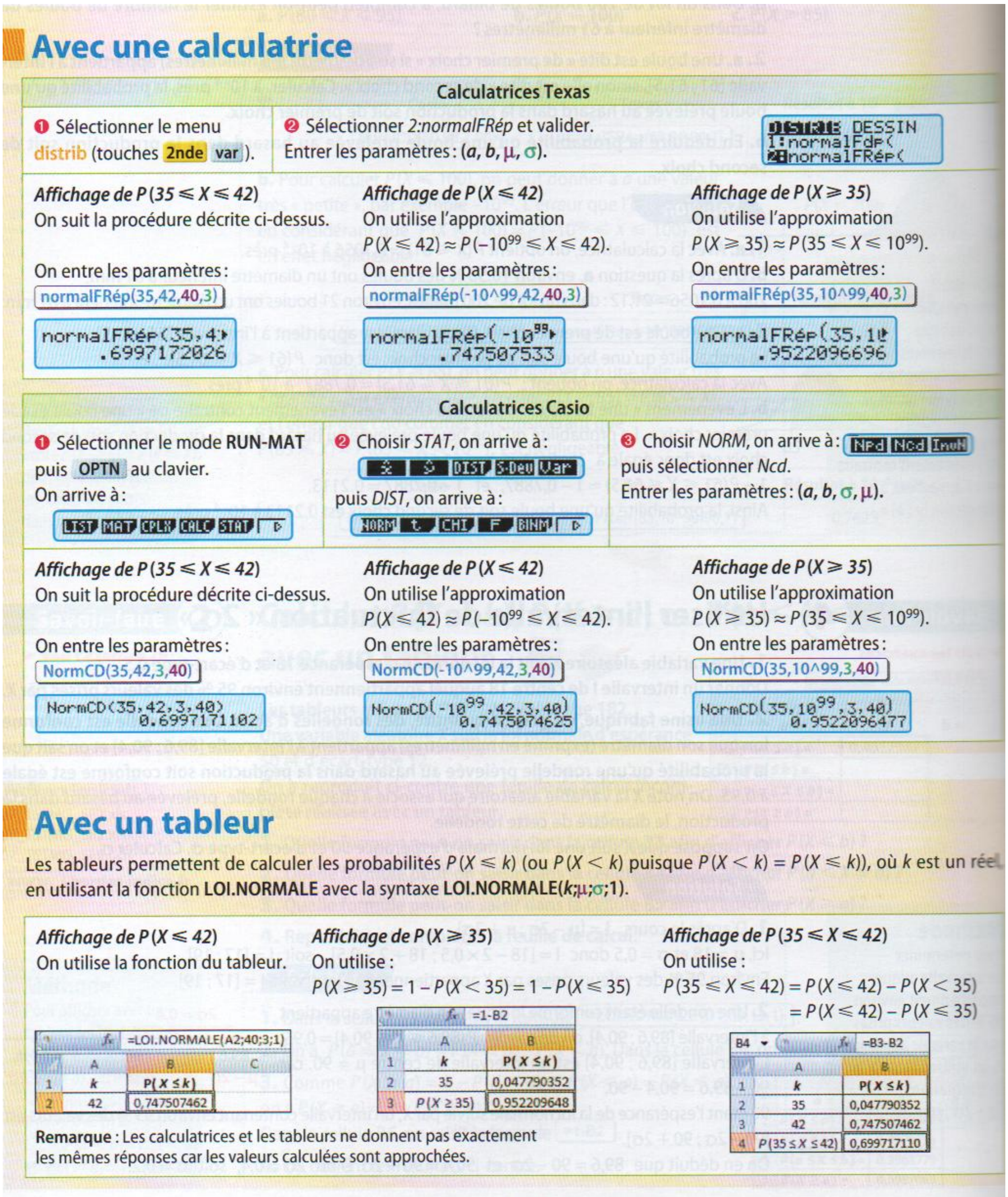
Exercices : 15, 16 page 140 ; 33 page 145 et 34 page 146[6] [Intervalle]

Références

[Intervalle] Collection Intervalle, Mathématiques, programme 2013, Term STMG, Nathan, 2013.

4, 5, 6

5. Échantillonnage, prise de décision.
6. Intervalle de confiance.



Avec une calculatrice

Calculatrices Texas

❶ Sélectionner le menu distrib (touches 2nde var).

❷ Sélectionner *2:normalFRép* et valider.
Entrer les paramètres : (*a*, *b*, μ, σ).

DISTRIB DESSIN
1:normalFdp(
2:normalFRép(

Affichage de $P(35 \leqslant X \leqslant 42)$	Affichage de $P(X \leqslant 42)$	Affichage de $P(X \geqslant 35)$
On suit la procédure décrite ci-dessus.	On utilise l'approximation $P(X \leqslant 42) \approx P(-10^{99} \leqslant X \leqslant 42)$.	On utilise l'approximation $P(X \geqslant 35) \approx P(35 \leqslant X \leqslant 10^{99})$.
On entre les paramètres : normalFRép(35,42,40,3)	On entre les paramètres : normalFRép(-10^99,42,40,3)	On entre les paramètres : normalFRép(35,10^99,40,3)
normalFRép(35,4▸ .6997172026	normalFRép(-10⁹⁹▸ .747507533	normalFRép(35,1▸ .9522096696

Calculatrices Casio

❶ Sélectionner le mode **RUN-MAT** puis OPTN au clavier.
On arrive à : LIST MAT CPLX CALC STAT ▷

❷ Choisir *STAT*, on arrive à : x̂ ŷ DIST S-Dev Var
puis *DIST*, on arrive à : NORM t CHI F BINM ▷

❸ Choisir *NORM*, on arrive à : Npd Ncd InvN
puis sélectionner *Ncd*.
Entrer les paramètres : (*a*, *b*, σ, μ).

Affichage de $P(35 \leqslant X \leqslant 42)$	Affichage de $P(X \leqslant 42)$	Affichage de $P(X \geqslant 35)$
On suit la procédure décrite ci-dessus.	On utilise l'approximation $P(X \leqslant 42) \approx P(-10^{99} \leqslant X \leqslant 42)$.	On utilise l'approximation $P(X \geqslant 35) \approx P(35 \leqslant X \leqslant 10^{99})$.
On entre les paramètres : NormCD(35,42,3,40)	On entre les paramètres : NormCD(-10^99,42,3,40)	On entre les paramètres : NormCD(35,10^99,3,40)
NormCD(35,42,3,40) 0.6997171102	NormCD(-10⁹⁹,42,3,40) 0.7475074625	NormCD(35,10⁹⁹,3,40) 0.9522096477

Avec un tableur

Les tableurs permettent de calculer les probabilités $P(X \leqslant k)$ (ou $P(X < k)$ puisque $P(X < k) = P(X \leqslant k)$), où k est un réel, en utilisant la fonction **LOI.NORMALE** avec la syntaxe **LOI.NORMALE(*k*;μ;σ;1)**.

Affichage de $P(X \leqslant 42)$
On utilise la fonction du tableur.

=LOI.NORMALE(A2;40;3;1)

	A	B	C
1	k	P(X ≤ k)	
2	42	0,747507462	

Affichage de $P(X \geqslant 35)$
On utilise :
$P(X \geqslant 35) = 1 - P(X < 35) = 1 - P(X \leqslant 35)$

=1-B2

	A	B
1	k	P(X ≤ k)
2	35	0,047790352
3	P(X ≥ 35)	0,952209648

Affichage de $P(35 \leqslant X \leqslant 42)$
On utilise :
$P(35 \leqslant X \leqslant 42) = P(X \leqslant 42) - P(X < 35)$
$= P(X \leqslant 42) - P(X \leqslant 35)$

B4 =B3-B2

	A	B
1	k	P(X ≤ k)
2	35	0,047790352
3	42	0,747507462
4	P(35 ≤ X ≤ 42)	0,699717110

Remarque : Les calculatrices et les tableurs ne donnent pas exactement les mêmes réponses car les valeurs calculées sont approchées.

Table 2: Utilisation de la calculatrice ou du tableur pour une loi normale