Table des matières

Avertissement 1
Prérequis indispensables ... 2
Comment traiter ... 3
Conseils généraux de travail ... 4

Rappels fondamentaux de mathématiques et de probabilités 5
Fractions ... 5
Identités remarquables ... 5
Puissances ... 6
Racine carrée d'un nombre algébrique ... 6
Analyse combinatoire ... 7
Probabilités ... 8

Testez vos connaissances ... 11
Réponses au test de connaissances ... 14

Séquence de travail n°1

Chapitre 1 – Rappels de statistique descriptive 17

I. Définitions ... 19
1. Les statistiques et la statistique ... 19
2. Population ... 19
3. Échantillon ... 20
4. Variable statistique ... 20
5. Présentation des données et représentation graphique ... 21

WWW.TALIB24.COM

II. Paramètres de position **27**
1. Mode 27
2. Médiane 27
3. Moyenne arithmétique 28

III. Paramètres de dispersion **29**
1. Étendue 29
2. Écart moyen 29
3. Variance 30
4. Écart type 30
5. Coefficient de variation : CV 31

IV. Applications **32**

Séquence de travail n°2

Chapitre 2 – Variables aléatoires et lois de probabilités **37**

I. Variables aléatoires **39**
1. Position du problème 39
2. Variable aléatoire discrète 40
3. Variable aléatoire continue 50

II. La loi binomiale **52**
1. Position du problème 52
2. Définition de la loi binomiale 53
3. Propriétés de la loi binomiale 54
4. Champ d'application de la loi binomiale 56
5. Application de la loi binomiale 57

III. La loi de poisson **58**
1. Définition de la loi de Poisson 58
2. Propriétés de la loi de Poisson et utilisation des tables 59
3. Approximation d'une loi binomiale par une loi de Poisson 60

Testez vos connaissances **65**
Réponses au test des connaissance **69**

Exercices supplémentaires **75**
Corrigés des exercices supplémentaires **78**

Séquence de travail n°3

Chapitre 3 – La loi normale ou loi de Gauss $\mathcal{N}(m, \sigma)$ **83**

I. Étude préliminaire **87**

II. Distribution de Gauss ou loi normale : $\mathcal{N}(m, \sigma)$ **88**
1. Définition 88
2. Loi Centrée réduite $\mathcal{N}(0, 1)$ 89
3. Intervalle de confiance d'une moyenne 93
4. A quoi sert la loi normale ? 95

III. Approximation d'une loi binomiale par une loi normale **98**

Testez vos connaissances **103**
Réponses au test des connaissances **106**

Exercices supplémentaires **111**
Corrigés des exercices supplémentaires **113**

WWW.TALIB24.COM

Table I 117
Fonction de répartition –
Table de la distribution normale réduite 117

Table II 118
Table de l'écart réduit 118

Annexe : la loi normale, reine des statistiques 119

Bibliographie 127

Avertissement

« Je n'enseigne pas, je raconte »
MONTAIGNE, Essais

Ce document se propose de vous fournir l'essentiel des connaissances qui vous permettront de mieux comprendre les concepts et les outils de la statistique. C'est un ouvrage d'initiation dont l'objectif principal est l'acquisition des techniques de base de la statistique ainsi que l'interprétation des résultats qui en découlent. Pour cela, les fondements mathématiques des théories exposées ne sont pas développés.
Afin de répondre aux difficultés que rencontrent les étudiants pour transposer les connaissances théoriques à l'application pratique, le document réunit l'essentiel des connaissances avec de nombreux exemples d'application illustrant les parties théoriques.

Les connaissances importantes, qu'il faut absolument garder à l'esprit, sont signalées en grisé dans le texte.

A la fin de chaque chapitre, un test d'évaluation des connaissances est proposé.

Attention ! l'évaluation des connaissances fait partie intégrante de votre cours. Des points particuliers peuvent être développés à ce niveau sans qu'ils aient été traités auparavant.

Certains exercices, fortement inspirés des épreuves d'examen du BTSA, présentent l'essentiel des réponses mais pas une rédaction complète qui vous sera exigée lors des épreuves d'examen.

Les connaissances s'enchaînent dans un ordre logique. Chaque nouvelle notion introduite suppose que d'autres notions sont connues.

WWW.TALIB24.COM

En commençant par découvrir ces nouvelles notions, notamment à l'aide des exemples proposés, vous pouvez rencontrer des difficultés dues à une mauvaise assimilation de notions précédentes.
Il faut donc systématiquement revenir en arrière et reprendre le cours mal assimilé. Ces allers et retours dans votre progression sont presque inévitables. Ne soyez donc pas découragés pour autant.
Vous verrez alors que, petit à petit, les nouvelles notions s'éclaircissent et se mémorisent de mieux en mieux.

Prérequis indispensables à l'étude de ce livret...

Dans ce livret, on rappelle d'abord quelques éléments de statistique descriptive. On traitera alors les notions de variables aléatoires, indispensables pour mieux assimiler les lois de probabilité, notamment la loi binomiale et la loi de Poisson. Compte tenu de l'importance de la loi normale dans les tests statistiques qui seront étudiés dans les autres livrets, un chapitre lui est consacré à part.

L'étude de ce livret suppose quelques connaissances de base concernant l'algèbre et les probabilités dont l'essentiel est rappelé en tout début de ce livret.
La réussite au premier test sur les probabilités est largement suffisante pour aborder ce livret sans problèmes majeurs.

Comment traiter un exercice de statistique ?

La rédaction d'un exercice d'un test d'évaluation, d'un devoir ou à une épreuve d'examen, doit être réalisée avec le plus grand soin.

- Faites d'abord une première lecture rapide de l'énoncé de manière à situer le problème posé en relation avec votre programme.
  - Quelles sont les données (nature de la variable, loi de probabilité, taille de l'échantillon, paramètres donnés...) ?
  - Que vous demande-t-on ?
  - Les questions sont-elles liées ?
  - Quelle table statistique utiliser ?
- Commencez alors par résoudre l'exercice sur du brouillon, question par question.
- A l'examen, on vous jugera à la démarche adoptée pour résoudre les exercices mais aussi **à la rédaction et à la présentation** du travail fourni, que beaucoup d'étudiants négligent en se contentant par exemple :
  - d' « appliquer » des formules sans expliquer les conditions d'applications,
  - d'aboutir par le calcul à des décisions « statistiques » mais sans une interprétation rigoureuse de leurs conclusions.

    Si vous rédigez, c'est pour être lu. Soignez vos copies. N'imposez pas à votre correcteur de vous « déchiffrer » . Il peut se lasser...

Vous risquez alors de perdre des points inutilement.

- Faites attention aux calculs numériques et aux unités. Les ordres de grandeurs doivent être respectés.
- Chaque résultat final d'une question doit être souligné proprement et suivi d'une petite conclusion.

WWW.TALIB24.COM

Conseils généraux de travail

Ce livret se présente sous forme de séquences de travail visant des objectifs pédagogiques formulés dès le départ. Les évaluations qui vous sont proposées à la fin des séquences visent à vérifier l'atteinte des objectifs visés par la séquence de travail proposée.

Pour cela, nous vous conseillons :

- de travailler aussi régulièrement que possible ;
- d'éloigner de votre vue tout ce qui peut vous distraire : magazines, journaux, radio, télé…
- d'avoir toujours sous la main une calculatrice, du brouillon, un crayon de papier et une gomme ;
- de vérifier, chaque fois que vous avez un doute, les calculs développés ;
- de traiter la totalité des exercices d'application proposés avant de passer à la séquence suivante ;
- d'établir une fiche de synthèse à la fin de chaque séquence de travail ; elle vous sera très utile pour la séquence suivante ;
- si vous avez la chance d'avoir un micro et de maîtriser EXCEL, n'hésitez pas à rentrer les données des exercices proposés et de faire exécuter les calculs et les graphiques par le logiciel ; cela vous permettra de faire des simulations en changeant les données pour « voir ce qui se passe ».

Tous les enseignants et pédagogues connaissent très bien la difficulté de rédiger un cours de statistique. Tous savent combien le traitement d'un problème de statistique est fortuit, si on fait l'impasse sur des concepts qui le sous-tendent. Ceux qui se référeront à ce document voudront bien l'utiliser avec indulgence et nous communiquer éventuellement, leurs remarques et suggestions. Nous les remercions.

Rappels fondamentaux de mathématiques et de probabilités

Fractions

- Pour tout $k \neq 0$, on a : $\frac{a}{b} = \frac{ka}{kb}$
- $\frac{a}{b} \times \frac{c}{d} = \frac{ac}{bd}$ avec b et $d \neq 0$
- $\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$ avec b et $d \neq 0$
- division par un nombre : $\frac{\frac{a}{b}}{c} = \frac{a}{bc}$
- division par une fraction : $\frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$

Identités remarquables

- $(a + b)^2 = a^2 + 2\,ab + b^2$
- $(a - b)^2 = a^2 - 2\,ab + b^2$
- $a^2 - b^2 = (a + b)(a - b)$
- $(a + b)^n = C_n^k\, a^k b^{n-k}$ (binôme de Newton)

C_n^k est le nombre de combinaisons de n objets pris k à k.

WWW.TALIB24.COM

Puissances

Par convention, on a : $a^1 = a$ et $a^0 = 1$

- $a^m a^n = a^{m+n}$
- $(abc)^m = a^m \times b^m \times c^m$
- $(a^m)^n = a^{mn}$
- $[\frac{a}{b}]^m = \frac{a^m}{b^m}$

Puissances entières relatives

- Calcul de $\frac{a^m}{b^m}$ avec $a \neq 0$, $m \in N$; $n \in N$

$$\frac{a^m}{a^n} = a^{m-n} \quad ; \quad a^{-n} = \frac{1}{a^n}$$

Puissances rationnelles

On peut écrire, si m et n sont deux entiers arithmétiques et $a \geq 0$:

$$\sqrt[m]{a^n} = a^{n/m}$$

Les propriétés des puissances entières relatives sont encore vraies.

Racine carrée d'un nombre algébrique

- Propriétés valables pour des nombres positifs ou nuls

- $\sqrt{a \times b \times c} = \sqrt{a} \times \sqrt{b} \times \sqrt{c}$
- Si $a \geq 0$ et $b > 0$ $\quad \sqrt{\frac{a}{b}} = \frac{\sqrt{a}}{\sqrt{b}}$

Attention ! $\sqrt{a+b} \neq \sqrt{a} + \sqrt{b}$

$\sqrt{a-b} \neq \sqrt{a} - \sqrt{b}$

Définition d'une intégrale

Le réel F (b) - F (a) est appelé intégrale de f de a à b. On note :

$$F(b) - F(a) = \int_a^b f(x)dx$$

Analyse combinatoire

Factorielle

Si n est un entier > 1, on appelle factorielle n et on note n ! Le produit des n premiers entiers naturels non nuls.
Si n = 1, on pose : 1 ! = 1 ; si n = 0, on pose : 0 ! = 1

Arrangements

On appelle arrangements de n objets (d'un ensemble E) pris p à p ($p \leq n$) tout sous-ensemble **ordonné** de E contenant p objets distincts. On démontre que leur nombre, noté A_n^p, est :

$$A_n^p = \frac{n!}{(n-p)!}$$

A_n^p est le nombre d'injections d'un ensemble à p éléments dans un ensemble à n éléments.

Permutations

C'est un cas particulier d'arrangement où n = p.

$$A_n^n = \frac{n!}{(n-n)!} = n!$$

WWW.TALIB24.COM

Combinaisons

On appelle combinaison de n objets (d'un ensemble E) pris p à p ($p \leq n$) tout sous-ensemble de E contenant p objets distincts. On démontre que leur nombre, noté C_n^p est :

$$C_n^p = \frac{n!}{p!\,(n-p)!}$$

Probabilités

Épreuve (ou expérience) aléatoire

C'est toute épreuve ou expérience dont l'issue n'est pas déterminable a priori.

Événement élémentaire

C'est toute issue possible d'une épreuve.

Univers ou ensemble fondamental

Ensemble, noté Ω, associé à une épreuve aléatoire. C'est l'ensemble des résultats possibles d'une épreuve aléatoire.

Événement

On appelle événement lié à une expérience aléatoire d'univers Ω, tout fait dont la réalisation dépend exclusivement de l'issue de cette expérience.

Lorsque les événements sont équiprobables, la probabilité d'un événement A est :

$$\text{Prob}(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

Si $\overline{A}$ est l'événement contraire de A :

$$\text{Prob}(\overline{A}) = 1 - \text{Prob}(A)$$

On appelle probabilité, définie sur Ω toute application P de $\mathcal{P}(\Omega)$, ensemble des parties de Ω vers $\Re^+$ vérifiant :

- Prob $(\Omega) = 1$, probabilité totale
- $A \cap B = \emptyset \Rightarrow \text{Prob}(A \cup B) = \text{Prob}(A) + \text{Prob}(B)$

On dit que A et B sont des événements incompatibles.

- $A \cap B \neq \emptyset \Rightarrow \text{Prob}(A \cup B) = \text{Prob}(A) + \text{Prob}(B) - \text{Prob}(A \cap B)$

Probabilités composées (ou conditionnelles)

Soit une expérience aléatoire d'univers Ω.

A est un événement de probabilité non nul.

B est un événement quelconque.

On appelle probabilité conditionnelle de B, sachant que A est réalisé, le nombre noté Prob (B/A) défini par :

$$\textbf{Prob (B/A)} = \frac{\textbf{Prob}(A \cap B)}{\textbf{Prob (A)}}$$

$\Rightarrow$ $\text{Prob}(A \cap B) = \text{Prob}(A) \times \text{Prob}(B/A)$

$\Rightarrow$ $\text{Prob}(A \cap B) = \text{Prob}(B) \times \text{Prob}(A/B)$

WWW.TALIB24.COM

Testez vos connaissances

Exercice 1

On considère une population de poulets parmi lesquels certains sont atteints d'un parasite.
Après traitement d'une partie des poulets, on remarque que :
• sur les 70 % de poulets non traités, 25 % sont atteints du parasite,
• sur les 30 % de poulets traités, 12,5 % sont encore atteints du parasite.
On prélève au hasard un poulet, on admet que tous les poulets ont la même probabilité d'être prélevés.
1. Quelle est la probabilité que ce poulet soit parasité ?
2. Sachant que ce poulet est parasité, quelle est la probabilité qu'il soit traité ?

NB : les résultats numériques seront donnés à 10^{-4} près.

Exercice 2

Dans une certaine population, il y a 45 % de fumeurs et 35 % de personnes atteintes de bronchite.
Sachant que parmi les fumeurs, il y a 65 % de bronchiteux, calculer la probabilité pour qu'une personne atteinte de bronchite soit fumeur.

WWW.TALIB24.COM

Exercice 3

Un élevage de porcs est constitué de 60 % de Landrace et de 40 % de Large White. On a pu établir que dans cet élevage :

- 7,5 % des Landrace sont atteints d'un parasite,
- 5 % des Large White sont atteints du même parasite.

On prélève au hasard un porc de l'élevage.

1. Calculer la probabilité qu'il soit atteint du parasite.

2. Sachant que le porc prélevé est parasité, calculer la probabilité qu'il soit de race Large White.

Exercice 4

Une cage contient 15 souris :

- 9 blanches dont 3 mâles et 6 femelles
- 6 grises dont 4 mâles et 2 femelles

Chaque souris a la même probabilité d'être prélevée.

1. Une personne retire une souris au hasard. Calculer la probabilité pour que :

a. la souris soit blanche

b. la souris soit mâle

c. la souris étant blanche, celle-ci soit un mâle

2. Une personne retire simultanément deux souris. Calculer la probabilité pour que :

a. les deux souris soient blanches

b. les deux souris soient de couleurs différentes

c. les deux souris soient du même sexe

Exercice 5

On considère une population composée de 45 % d'hommes et 55 % de femmes. On suppose que 4 % des hommes et 0,5 % des femmes sont daltoniens. On choisit au hasard une personne dans cette population.

1. Quelle est la probabilité pour que cette personne soit daltonienne ?

2. Sachant que cette personne est daltonienne, quelle est la probabilité pour que ce soit un homme ?

Vérifiez vos réponses à la page suivante

WWW.TALIB24.COM

Réponses au test

Exercice 1

1. Soient les événements :

T : le poulet est traité.

Pa : le poulet est parasité.

$Pa = (T \cap Pa) \cup (\overline{T} \cap Pa)$ les événements entre parenthèses $(T \cap Pa)$ $(\overline{T} \cap Pa)$ sont incompatibles

d'où $P(Pa) = P(T \cap Pa) + P(\overline{T} \cap Pa)$

$P(T \cap Pa) = P(T) \times P(Pa / T) = 0,30 \times 0,125 = 0,0375$

$P(\overline{T} \cap Pa) = P(\overline{T}) \times P(Pa / \overline{T}) = 0,70 \times 0,25 = 0,1750$

d'où $P(Pa) = 0,2125$

2. On cherche $P(T / Pa) = \dfrac{P(T \cap Pa)}{P(Pa)} = \dfrac{0,0375}{0,2125} = 0,1765$

Exercice 2

Soient les événements :

F la personne est fumeur

B la personne est bronchiteuse

On cherche $P(F / B) = \dfrac{P(F \cap B)}{P(B)}$

$P(F \cap B) = P(F) \times P(B / F) = 0,45 \times 0,65 = 0,2925$

d'où $P(F / B) = \dfrac{0,2925}{0,35} = 0,8357$

Exercice 3

Soit les événements :

L : le porc est Landrace.

Pa : le porc est parasité.

1. $Pa = (L \cap Pa) \cup (\overline{L} \cap Pa)$

$(L \cap Pa)$ et $(\overline{L} \cap Pa)$ sont incompatibles d'où

$P(Pa) = P(L \cap Pa) + (\overline{L} \cap Pa)$

$P(L \cap Pa) = P(L) \times P(Pa / L) = 0,60 \times 0,075 = 0,045$

$P(\overline{L} \cap Pa) = P(\overline{L}) \times P(Pa / \overline{L}) = 0,40 \times 0,05 = 0,02$

d'où $P(Pa) = 0,065$

2. On cherche $P(\overline{L} / Pa) = \dfrac{P(\overline{L} \cap Pa)}{P(Pa)} = \dfrac{0,02}{0,065} = 0,3077$

Exercice 4

1.

a. P (souris est blanche) = $\dfrac{\text{nombre de cas favorables à l'événement}}{\text{nombre de cas possibles}}$

$P(Bl) = \dfrac{9}{15}$

b. P(la souris est un mâle) = 7/15 = P(♂)

c. $P(♂ / Bl) = \dfrac{P(♂ \cap Bl)}{P(Bl)} = \dfrac{\frac{3}{15}}{\frac{9}{15}} = \dfrac{3}{9} = \dfrac{1}{3}$

2.

WWW.TALIB24.COM

a. P (les 2 souris sont blanches) = $\dfrac{C_9^2 \times C_6^0}{C_{15}^2} = \dfrac{36}{105} = 0{,}3429$

b. P (les 2 souris sont de couleurs différentes) : $\dfrac{C_9^1 \times C_6^1}{C_{15}^2} = \dfrac{54}{105} = 0{,}5143$

c. P (les 2 souris sont du même sexe) = P (2 souris sont donc ♂ ou ♀)

$= P(2\,♂) + P(2\,♀) = \dfrac{C_7^2 + C_8^2}{C_{15}^2} = \dfrac{21+28}{105} = 0{,}4667$

Exercice 5

1. Soit les événements :

H : la personne est un homme

F : la personne est une femme

D : la personne est daltonienne.

$D = (H \cap D) \cup (F \cap D)$

$(H \cap D)$ et $(F \cap D)$ sont des événements incompatibles, d'où

$P(D) = P(D \cap H) + P(D \cap F)$

$P(H \cap D) = P(H) \times P(D/H) = 0{,}45 \times 0{,}04 = 0{,}018$

$P(F \cap D) = P(F) \times P(D/F) = 0{,}55 \times 0{,}005 = 0{,}0028$

d'où $P(D) = 0{,}0208$

2. On cherche

$$P(H/D) = \frac{P(H \cap D)}{P(D)} = \frac{0{,}018}{0{,}0208} = 0{,}8654$$

SÉQUENCE DE TRAVAIL N°1

Rappels de statistique descriptive

Objectifs pédagogiques

A la fin de cette séquence, mais étape par étape, vous devriez être capable de :

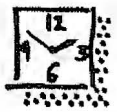
1. différencier une variable continue d'une variable discrète ;

2. distinguer une population d'un échantillon ;

3. représenter graphiquement des données relatives :
- à une variable continue,
- à une variable discrète,

4. définir et de calculer une moyenne arithmétique, une variance et un écart type d'un échantillon ;

5. calculer et d'interpréter un coefficient de variation.



3 h

WWW.TALIB24.COM

I. Définitions

1. Les statistiques et la statistique

Les **statistiques** sont des collections de nombres présentées sous forme de tableaux ou de graphiques groupant des observations relatives à un phénomène considéré.

La **statistique** est une méthode de raisonnement permettant d'interpréter le genre de données très particulières, qu'on rencontre notamment dans les sciences de la vie, dont le caractère essentiel est **la variabilité.**

- **La statistique descriptive** est un ensemble de méthodes qui permettent d'ordonner et de classer les données, de les réduire ensuite à un nombre limité de paramètres caractéristiques (moyenne, variance, écart type...) susceptible de décrire la distribution du caractère étudié dans une population donnée.
- **La statistique inductive** est plus ambitieuse car elle cherche les principes permettant de déduire des résultats obtenus sur un échantillon limité, une généralisation à l'ensemble de la population d'où est extrait cet échantillon et qui est généralement inaccessible à l'enquête ou à la mesure. On est alors conduit à formuler des hypothèses dont on vérifie la validité à l'aide de certaines épreuves ou **tests statistiques.**

2. Population

L'objet de toute étude statistique est de formuler des lois valables pour un ensemble d'êtres ou d'éléments, auquel on donne le nom de population. Cette dernière est l'ensemble de toutes les observations possibles concernant le caractère étudié.

WWW.TALIB24.COM

Une population peut être de nature très variée : êtres humains, animaux, plantes, bactéries etc.
Il est évidemment peu commode d'étudier tout cet ensemble à cause du nombre très élevé du nombre d'observations (parfois infini).

3. Échantillon

L'étude d'un caractère porte en général sur une partie restreinte de la population appelée alors échantillon dont l'**effectif n est fini.** Le choix de cet échantillon sera fait au hasard.

On a coutume de distinguer :
- **les grands échantillons, lorsque $n \geq 30$**
- **les petits échantillons, lorsque $n < 30$**

4. Variable statistique

Le caractère sur lequel porte l'étude statistique est appelée variable statistique. Cette dernière peut être :
- **qualitative :** couleur des yeux, réaction à un vaccin...
- **quantitative :** l'intensité du caractère peut être alors mesurée.

Une variable quantitative peut être :
- **continue** : si l'intensité du caractère étudié peut prendre toutes les valeurs d'un intervalle fini ou infini ; exemple : longueur des tiges d'une plante, poids de nouveau-nés...
- **discontinue :** (on dit aussi discrète) ; exemple : nombre d'étudiants par classe, nombre d'enfants par couple...

5. Présentation des données et représentation graphique

A. Présentation des données

◆ *Cas d'un caractère quantitatif discontinu*

– Soit un échantillon de taille n, c'est-à-dire comportant n éléments numérotés par exemple de 1 à n.

– Si X est la variable sur lequel porte l'étude, alors X peut prendre les valeurs :

$$x_1 ; x_2 ; \ldots\ldots ; x_i ; \ldots\ldots\ldots ; x_n$$

– L'ensemble des valeurs de X constitue une **série statistique** définie par :
- son **étendue,** c'est-à-dire l'écart qui sépare la plus grande de la plus petite valeur ;
- son effectif total, c'est-à- dire la somme de tous les x_i

Si la valeur x_i du caractère se répète un nombre de fois n_i dans la série statistique, n_i est alors une **répétition** de la valeur x_i.
n_i est appelé **fréquence absolue de x_i.**
On peut donc écrire :

$$n = \Sigma\, n_i$$

– Le nombre $f_i = n_i / n$ est appelé **fréquence relative de x_i.**
Bien évidemment, la somme des fréquences relatives est égale à 1 :

$$\Sigma\, f_i = 1$$

WWW.TALIB24.COM

Exemple

x_i caractère	n_i fréquence absolue	f_i fréquence relative
5	4	0,0125
10	19	0,0594
15	33	0,1031
20	63	0,1969
25	82	0,2563
30	56	0,1750
35	38	0,1188
40	20	0,0625
50	5	0,0156
Σ	n = 320	1

◆ ***Cas d'un caractère quantitatif continu***

Un caractère continu a un nombre de valeurs distinctes **théoriquement infini**. En réalité, et pour des raisons pratiques, nous ne pourrons discerner qu'un nombre fini de valeurs distinctes qui forment un ensemble discontinu de valeurs.

Ainsi, on constitue des **classes** en divisant l'étendue de la série en un certain nombre d'intervalles partiels, chacune des classes groupant ensemble, les mesures relatives à un même intervalle.

Chaque classe contiendra toutes les valeurs supérieures ou égales à sa limite inférieure, mais strictement inférieures à sa limite supérieure.

Le centre de classe représente la « moyenne » de la classe et l'étendue de la classe est l'écart entre les deux limites de classe.

Exemple

Taille des enfants d'une école (en cm)

- la classe **[100 – 105[** du tableau ci-dessous contiendra l'ensemble des valeurs de x comprises entre 100 (inclus) et 105 (exclu) ;
- la classe suivante **[105 – 110[** contiendra l'ensemble des valeurs de x comprises entre 105 (inclus) et 110 (exclu) ;

etc.

En pratique les classes sont d'étendue égale, mais cela n'est pas indispensable

Classes	Limites de la classe (cm)	x_i caractère centre de classe (cm)	n_i fréquence absolue	f_i fréquence relative
1	[100 – 105[	102,5	4	0,0125
2	[105 – 110[	107,5	19	0,0594
3	[110 – 115[	112,5	33	0,1031
4	[115 – 120[	117,5	63	0,1969
5	[120 – 125[	122,5	82	0,2563
6	[125 – 130[	127,5	56	0,1750
7	[130 – 135[	132,5	38	0,1188
8	[135 – 140[	137,5	20	0,0625
9	[140 – 145[	142,5	5	0,0156
		Σ	n = 320	1

WWW.TALIB24.COM

◆ ***Cas d'un caractère qualitatif***

Les résultats d'une enquête relative à un caractère qualitatif, sont groupés en autant de catégories ou classes qu'il existe de modalités pour le caractère étudié.

A chaque classe est alors associé son effectif n_i, ou **fréquence absolue** et sa **fréquence relative** $f_i = n_i / n$

Exemple

Soient deux gènes codominants A et a. Le croisement, Aa × Aa donnera théoriquement, en deuxième génération :

25 % du phénotype [A] ; 25 % du phénotype [a] ; 50 % du phénotype [Aa]

Sur 150 individus, on a obtenu :

Phénotype	n_i fréquence absolue	f_i fréquence relative
[A]	40	0,266
[Aa]	82	0,546
[a]	38	0,253
	150	1,000

B. Représentation graphique

• Cas d'un caractère discontinu : **diagramme en bâtons**

C'est un ensemble de segments verticaux ayant pour origine les points d'abscisse x_i et pour hauteur une longueur proportionnelle :

- soit à l'effectif correspondant ;
- soit à la fréquence f_i correspondante.

◆ ***Polygone des fréquences***

On obtient le polygone des fréquences en joignant par des segments de droite les extrémités des bâtons.

◆ ***Diagramme cumulatif***

L'effectif cumulé jusqu'à la $i^{ème}$ valeur x_i du caractère est la somme $n_1 + n_2 + \ldots\ldots\ldots\ldots + n_i$ des effectifs partiels obtenue pour les i premières valeurs du caractère.

La fréquence relative cumulée jusqu'à la $i^{ème}$ valeur x_i du caractère est la somme $f_1 + f_2 + \ldots\ldots\ldots\ldots + f_i$

Exemple

x_i	n_i	Effectif cumulé	Fréquence relative	Fréquence relative cumulée
110	4	4	0,0125	0,0125
115	19	23	0,0594	0,0719
120	33	56	0,1031	0,1750
125	63	119	0,1969	0,3719
130	82	201	0,2563	0,6281
135	56	257	0,1750	0,8031
140	38	295	0,1188	0,9219
145	20	315	0,0625	0,9844
150	5	320	0,0156	1,0000
Somme	320		1	

Dans le diagramme cumulatif, les bâtons ont des longueurs proportionnelles aux effectifs cumulés.

WWW.TALIB24.COM

• Cas d'un caractère continu : **histogramme**

Le diagramme en bâtons est remplacé par un histogramme. En effet, la représentation précédente n'est plus possible puisque, à l'intérieur d'une même classe, le nombre de valeurs discernables de x est très grand.

L'aire de l'histogramme est proportionnelle à l'effectif total.

II. Paramètres de position

1. Mode

Le mode d'une série statistique est la valeur de la variable qui a l'effectif le plus élevé. Le mode peut ne pas exister et s'il existe, il peut ne pas être unique (série monomodale ou plurimodale).

Exemple : 2, 2, 5, 8, 9, 9, 9, 10, 12, 13 a pour mode 9
3, 4, 5, 6, 7, 8, 9, 10 n'a pas de mode

2. Médiane

La médiane d'un ensemble de nombres rangés par ordre de grandeur non décroissante est la valeur du milieu ou la moyenne arithmétique des valeurs centrales.

Exemple : 3 8 14 15 18 22 23
médiane = 15

3 8 14 15 18 22

médiane = $\frac{14+15}{2}$

La médiane est déterminée par le nombre d'observations, mais non par la valeur de celles-ci. Ainsi, les valeurs extrêmes, grandes ou petites, n'affectent pas la médiane.

WWW.TALIB24.COM

3. Moyenne arithmétique

La moyenne arithmétique $\overline{x}$ d'une série de mesure $x_1, x_2, \ldots x_n$ d'une même grandeur x, est définie par :

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum n_i x_i}{n}$$

C'est la mesure la plus familière et la plus couramment utilisée. Elle est influencée par la valeur de **toutes** les observations. Sa valeur peut être biaisée par quelques valeurs extrêmes.

◆ ***Propriétés de la moyenne arithmétique***

La somme algébrique des écarts à la moyenne est nulle.

Si, $\overline{x} = \dfrac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$

alors, $x_1 + x_2 + x_3 + \ldots + x_n - n\overline{x} = 0$

ou $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$

III. Paramètres de dispersion

1. Étendue

C'est la différence entre le plus grand et le plus petit nombre d'une série statistique.

Exemple : l'étendue de l'ensemble des nombres :

2 3 3 5 5 5 8 10 12 est $12 - 2 = 10$

2. Écart moyen

L'écart moyen (EM) d'un ensemble de nombres $x_1, x_2, \ldots x_n$ est défini par :

$$EM = \frac{\sum |x_i - \overline{x}|}{n}$$

$\overline{x}$ étant la moyenne arithmétique.

Exemple

Calcul de l'EM des nombres 2, 3, 6, 8, 11

Valeur moyenne : $\overline{x} = \dfrac{2+3+6+8+11}{5} = 6$

$$EM = \frac{|2-6| + |3-6| + \ldots\ldots |11-6|}{5} = 2{,}8$$

Les principales caractéristiques de l'écart moyen sont les suivantes :

- il accorde le même poids à tous les écarts ;
- il peut être calculé aussi bien à partir de la médiane qu'à partir de la moyenne ;
- c'est une mesure dont le calcul et l'interprétation sont simples.

WWW.TALIB24.COM

3. Variance

C'est la moyenne arithmétique des carrés des écarts par rapport à la moyenne d'une série statistique.
On la note V (x) .

$$V(x) = \frac{\sum n_i (x_i - \bar{x})^2}{n}$$

L'expression $\sum n_i (x_i - \bar{x})^2$ s'appelle somme des carrés des écarts relatifs à la moyenne (**SCE**). L'expression de la variance s'écrit alors :

$$V(x) = \frac{SCE}{n}$$

On montre aussi que,

$$V(x) = \left[\frac{\sum n_i x_i^2}{n}\right] - \bar{x}^2$$

C'est cette dernière formule qui est souvent utilisée. Son avantage réside dans le fait que la moyenne $\bar{x}$ n'intervient qu'une seule fois dans le calcul final.

La variance V(x) peut être désignée par σ^2 (x) ou tout simplement σ^2

4. Écart type

On définit l'écart type $\sigma(x)$ d'un ensemble N de nombres $x_1, x_2 ..... x_n$ par :

$$\sigma(x) = \sqrt{V(x)}$$

C'est donc la racine carrée de la variance.

Certaines calculatrices scientifiques ont des touches notées σ ou σ_{n-1} qui permettent de calculer :

$$\sigma \text{ (ou } \sigma_{n-1}) = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{n-1}}$$

En fait cette formule n'est valable que pour des échantillons extraits d'une population (cf. chapitre sur l'estimation)

Caractéristiques de l'écart type :

- c'est la mesure de dispersion la plus fréquemment utilisée ;
- la valeur de chacune des observations affecte la valeur de l'écart type. Un changement de valeur d'une observation aura pour effet de changer la valeur de l'écart type ;
- sa valeur peut être fortement influencée par quelques valeurs extrêmes. A l'instar de la moyenne, l'écart type tend à être moins représentatif dans les distributions fortement dissymétriques ;
- on ne peut le calculer dans des distributions à classes ouvertes qu'à l'aide d'informations additionnelles.

5. Coefficient de variation : CV

C'est le rapport entre l'écart type et la moyenne :

$$CV = \frac{\sigma}{x} \times 100$$

Le CV est exprimé en pourcentage (%). Il résume la dispersion des résultats et est utilisé surtout dans les essais d'agronomie. On considère qu'un essai est fiable si le CV≤ 12 - 15 %

WWW.TALIB24.COM

L'étendue, l'écart moyen, la variance et l'écart type sont des mesures de dispersion **absolue.** Elles s'expriment dans les mêmes unités que les observations originales. Ces mesures ne permettent pas de comparer deux dispersions portant sur deux observations différentes par leurs unités. Le coefficient de variation, mesure de dispersion **relative,** permet une telle comparaison.

Exemple

	Distribution des tailles (cm)	Distribution des poids (kg)
Moyenne	160 cm	70 kg
Écart type	20 cm	5 kg
Coefficient de variation	(20/160) .100 = 12,5 %	(5/70) .100 = 7,14 %

Dans cet exemple, la dispersion relative est plus grande pour la variable « taille » que pour la variable « poids » ; autrement dit la distribution des poids est plus homogène que la distribution des tailles.

IV. Applications

Exercice

La série suivante représente la croissance journalière (en g/j) d'une série d'agneaux pendant le premier mois (GMQ 10-30 j).

Croissance des agneaux, x (g/j)	110	115	120	125	130	135	140	145	150
Effectif, n	4	19	33	63	82	56	38	20	5

1. Calculer la moyenne, la variance et l'écart type de cette série.
2. Établir l'histogramme et le polygone de fréquence de cette série.
3. Établir l'histogramme des effectifs cumulés croissants ainsi que le polygone des fréquences cumulées relatives croissantes.

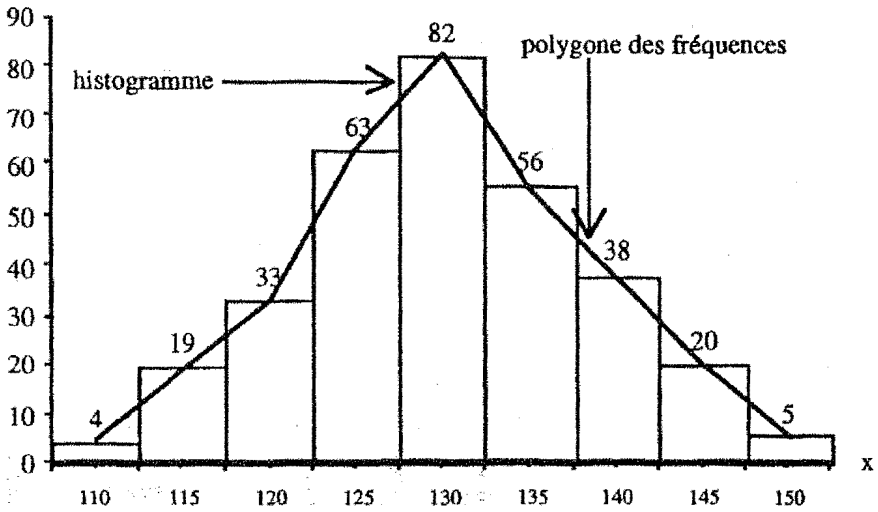
Corrigé

1. Calcul de la moyenne, la variance et l'écart type de la série.
Le tableau ci-dessous résume la présentation des résultats et les étapes du calcul.

x_i	n_i	$n_i x_i$	x_i^2	$n_i x_i^2$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
110	4	440	12100	48400	406,3	1625,1
115	19	2185	13225	251275	229,7	4364,5
120	33	3960	14400	475200	103,1	3403,9
125	63	7875	15625	984375	26,6	1675,0
130	82	10660	16900	1385800	0,0	2,0
135	56	7560	18225	1020600	23,5	1313,9
140	38	5320	19600	744800	96,9	3682,2
145	20	2900	21025	420500	220,3	4406,7
150	5	750	22500	112500	393,8	1968,9
Somme	**$\Sigma n_i = n = 320$**	**41650**	**153600**	**5443450**	**1500,2**	**22442,2**
Moyenne, $\bar{x}$	130,16 g	$\bar{x} = \frac{\sum x_i}{n}$				
Variance, V(x)	70,13 g²	$V(x) = \frac{\sum n_i(x_i - \bar{x})^2}{n} = \left[\frac{\sum n_i x_i^2}{n}\right] - \bar{x}^2$				
Écart type, σ(x)	8,37 g	$\sigma(x) = \sqrt{V(x)}$				

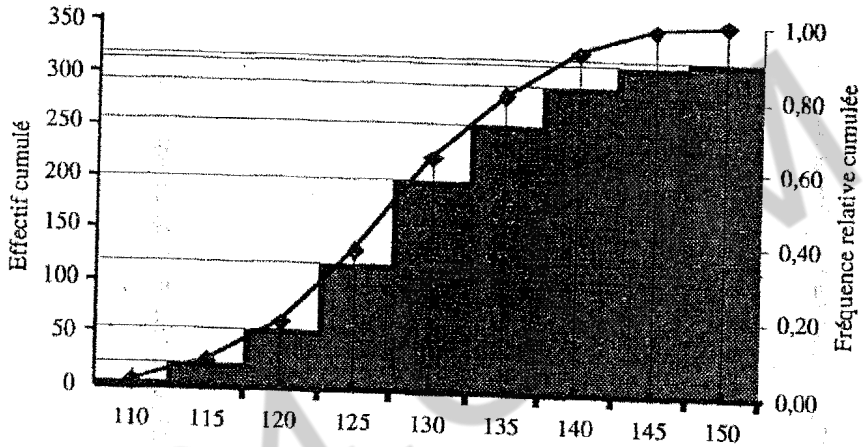
WWW.TALIB24.COM

2. Représentation graphique



3. Histogramme des effectifs cumulés croissants
et polygone des fréquences relatives cumulées croissantes

x_i	n_i	Effectif cumulé	Fréquence relative	Fréquence relative cumulée
110	4	4	0,0125	0,0125
115	19	23	0,0594	0,0719
120	33	56	0,1031	0,1750
125	63	119	0,1969	0,3719
130	82	201	0,2563	0,6281
135	56	257	0,1750	0,8031
140	38	295	0,1188	0,9219
145	20	315	0,0625	0,9844
150	5	320	0,0156	1,0000
Somme	320		1	



WWW.TALIB24.COM

SÉQUENCE DE TRAVAIL N°2

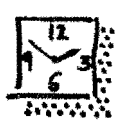
Variables aléatoires et lois de probabilités

2

Objectifs pédagogiques

A la fin de cette séquence, mais étape par étape, vous devriez être capable :

1. de définir une variable aléatoire et de distinguer une variable aléatoire continue d'une variable aléatoire discrète ;
2. de définir une loi de probabilité ;
3. de distinguer une fonction de densité de probabilité d'une fonction de répartition ;
4. d'appliquer une loi binomiale et d'en décrire les paramètres ;
5. d'appliquer une loi de Poisson et d'en décrire le paramètre.



6 h

WWW.TALIB24.COM

I. Variables aléatoires

1. Position du problème

Il arrive souvent que le résultat d'une épreuve puisse s'exprimer par un nombre. Par exemple, le numéro sorti en jetant un dé, le nombre de « piles » obtenues en jetant 10 fois une pièce de monnaie, le nombre de coups de téléphone reçus par un établissement en une heure...

Mais il peut arriver aussi que les événements élémentaires d'une épreuve ne s'expriment pas par des nombres. Conventionnellement, on établit une correspondance entre ces événements et un ensemble des nombres. Par exemple, l'ensemble E des événements élémentaires relatifs à l'épreuve « lancer une pièce de monnaie » s'écrit :

$$E = \{P, F\}$$

On fera correspondre alors le nombre 1 au résultat « pile » et le nombre 0 au résultat « face ». L'ensemble E peut se remplacer par le nouvel ensemble $\{0, 1\}$.

Dans ces conditions, on appellera variable aléatoire une **application** de l'ensemble des événements élémentaires sur R (ensemble des nombres réels).

Une **variable aléatoire** X (que l'on notera parfois v.a. X) est donc un nombre associé à une expérience aléatoire ou une épreuve E et servant à caractériser le résultat de cette expérience.

Cette variable peut être :

- **discrète** ou **discontinue** ou **stochastique** si elle prend des valeurs séparées.
  Exemples : nombre d'enfants par famille ; nombre de lapereaux dans une portée ; nombre de lactations réalisées par une vache...

WWW.TALIB24.COM

- **continue** si elle peut prendre toutes les valeurs réelles de l'intervalle $(-\infty, +\infty)$.
  Exemples : taux de glycémie; production laitière d'une vache ; taille ou poids d'un animal…

2. Variable aléatoire discrète

A. Fonction de distribution

Exemple

On lance une paire de dés. On considère la **somme des points obtenus** après chaque jet. Cette expérience peut donner lieu à 11 résultats distincts résumés dans le tableau ci-dessous.

dé 1 / dé 2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

On peut associer à cette expérience une variable X pouvant prendre chacune des valeurs

$$x_i = 2, 3, 4 \ldots\ldots 12$$

avec les probabilités correspondantes $p_i = 1/36, 2/36, 3/36 \ldots\ldots 1/36$.

x_i	2	3	4	5	6	7	8	9	10	11	12	Σ
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	36/36

Pour obtenir 5, il faut : 4 et 1 ; 1 et 4 ; 2 et 3 ; 3 et 2
soit $1/36 \times 4 = 4/36 = 1/9$

- L'ensemble des valeurs possibles de la variable aléatoire et les probabilités correspondantes constituent une **distribution de probabilité**.
- On dit que X est une **variable aléatoire** et l'ensemble des couples (x_i, p_i) constitue, par définition, la **loi de probabilité** de la variable aléatoire X. Cette loi peut être représentée par un **diagramme en bâtons** (ci-dessous).

Représentation graphique de la distribution de probabilité

On appelle donc **loi de probabilité** de la v.a. X, la fonction P qui, à chaque x_i associe :

$$\text{Prob}\,(X = x_i) = \text{Prob}\,(x_i)$$

Si la v.a. X peut prendre n valeurs différentes x_i, on aura :

$$\sum_{i=1}^{i=n} \text{Prob}\,(x_i) = 1$$

WWW.TALIB24.COM

Lorsqu'on associe à chaque valeur de xi sa probabilité de réalisation, on réalise bien une application de X vers [0, 1]. Cette application est notée f(x)

$$f(x) = \text{Prob}\ (X = x)$$

x étant une valeur numérique particulière de X.

f(x) est la fonction de distribution. Elle est notée par un f minuscule.

B. Fonction de répartition

En **cumulant** les probabilités, on obtient une **distribution de probabilités cumulées.** La fonction associée à une telle distribution est appelée **fonction de répartition.** Elle est notée F(X).

La fonction de répartition d'une variable aléatoire discrète X, est donc la fonction F qui, à chaque valeur x de X, associe F(x) telle que :

$$F : \quad x \longrightarrow F(x) = \text{Prob}(X \leq x)$$

x étant une valeur numérique particulière de X.

En pratique, c'est la probabilité pour que X prenne une valeur **inférieure ou égale** à une valeur donnée x.

Certains auteurs confondent souvent x et X. En fait, il faut bien distinguer la variable aléatoire X – notée toujours par une majuscule – de la valeur numérique, x, que peut prendre cette variable et qui est notée toujours par une minuscule.

Exemple

Reprenons l'exemple du paragraphe I.2. : jet de deux dés dont on s'intéresse à la somme des points obtenus.

Appelons S la somme des points obtenus et considérons les cas suivants :

- Si $S \leq 3$, on perd 2 F.
- Si $3 < S \leq 6$, on perd 1 F.
- Si $6 < S \leq 9$, on ne gagne rien mais on ne perd rien non plus.
- Si $9 < S \leq 12$, on gagne 2 F.

Le tableau suivant résume l'ensemble des résultats possibles lors du jet de ces deux dés, c'est-à-dire les valeurs que peut prendre S.

résultat du jet dé 1 → dé 2 ↓	*1*	*2*	*3*	*4*	*5*	*6*
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Si, la variable aléatoire X représente le « gain » (positif ou négatif) la loi de probabilité de X sera alors :

X	–2	–1	0	2
Prob (X)	3/36	12/36	15/36	6/36

- Fonction de répartition

Pour $x \in]-\infty\ ;\ -2[\Rightarrow F(X) = 0$

Pour $x \in [-2\ ;\ -1[\Rightarrow F(X) = 3/36 = 1/12$

Pour $x \in [-1\ ;\ 0[\Rightarrow F(X) = 15/36 = 5/12$

Pour $x \in [0\ ;\ 2[\Rightarrow F(X) = 30/36 = 5/6$

Pour $x \in [2\ ;\ +\infty[\Rightarrow F(X) = 36/36 = 1$

La représentation graphique de la fonction de répartition se fait à l'aide d'un **diagramme en escalier** (cf. § Application, plus loin).

WWW.TALIB24.COM

C. Espérance mathématique

Partons de deux exemples :

Exemple 1

Considérons l'épreuve consistant à jeter une pièce de monnaie. On aura deux événements possibles : pile ou face. Chaque événement a une probabilité de 0,5 d'apparaître.

- Si l'événement « pile » apparaît, on gagne 10 F.
- Si l'événement « face » apparaît, on perd 20 F.

Événements	Face	Pile
X	– 20 F	+ 10 F
Prob(X = x)	0,5	0,5
Prob(X ≤ x)	0,5	0,5 + 0,5 = 1

Exemple 2

Considérons l'épreuve d'un jet de dé équilibré à 6 faces.

Les événements possibles sont {1,2,3,4,5,6}. Chaque événement a une probabilité de 1/6 d'apparaître.

- Si la face affiche un nombre premier (1, 2, 3, 5), on gagne 6 F sinon on perd 12 F.

Événements	4, 6	1, 2, 3, 5
X	-12 F	+ 6 F
Prob(X = x)	2/6	4/6
Prob(X ≤ x)	2/6	1

Dans les deux exemples, la variable X = « gain » ne peut prendre que deux valeurs possibles qui sont + 10 F ou –20 F dans l'exemple 1 et + 6 F et – 12 F dans l'exemple 2. Il n'existe pas de gains possibles entre deux valeurs consécutives. **X est bien une variable aléatoire discrète** ou un **aléa discret.**

Définition de E(X)

Soit X une v.a. discrète pouvant prendre les valeurs x_i (i = 1, 2, 3, ...n) avec les probabilités p_i.

E(X) correspond à l'espérance mathématique et se calcule à partir de l'expression :

$$m = E(X) = \sum_{i=1}^{n} p_i x_i$$

Dans l'exemple 1,

$$E(X) = 0{,}5 \times (+10) + 0{,}5 \times (-20) = -5 \text{ F}$$

Dans l'exemple 2,

$$E(X) = 2/6 \times (-12) + 4/6 \times (+6) = 0 \text{ F}$$

- Dans l'exemple 1, on perd en moyenne plus qu'on ne gagne. C'est le cas de la quasi totalité des jeux de la Loterie Nationale.
- Dans l'exemple 2, il y a équilibre à long terme entre les gains et les pertes.
- Si E(X) est positif, il s'agit d'un jeu utopique ; on gagne en moyenne plus qu'on ne perd.

Propriétés de E(X)

Ces propriétés sont très utiles pour le calcul numérique.

On démontre que :

- E(X + Y) = E(X) + E (Y)
- E(k.X) = k E(X) et k ∈ R
- E(X.Y) = E(X) . E(Y) si les deux variables X et Y sont indépendantes.
- E(X.Y) = E(X) . E(Y) + Cov(X ; Y) si les deux variables X et Y sont dépendantes et où :

$$Cov(X ; Y) = \frac{1}{n} \sum_{i=1}^{n} [x_i - E(x)]\,[y_i - E(Y)]$$

La covariance (Cov) de deux variables aléatoires indépendantes est nulle.

WWW.TALIB24.COM

• Si l'on majore ou l'on minore d'une même valeur k les valeurs prises par une variable aléatoire X, son espérance mathématique est majorée ou minorée de la même valeur k.

$$E(X + k) = E(X) + k$$

• $E(X - E(X)) = 0$; $X - E(X)$ est une variable centrée.

D. Variance

Définition de V(X)

$$V(X) = \sum_{i=1}^{n} p_i (x_i - m)^2$$

Propriétés de V(X)

Ces propriétés sont très utiles pour le calcul numérique.

• Si l'on ajoute une constante k à une variable aléatoire X, la variance est inchangée

$$V(X + k) = V(X)$$

• Si l'on multiplie par une constante k la variable aléatoire X, sa variance est multipliée par k^2.

$$V(kX) = k^2 V(X)$$

• On démontre que :

$$V(X) = E(X^2) - [E(X)]^2$$

et

$$V(X) = E\left[[X - E(X)]^2\right]$$

• Si X et Y sont deux variables aléatoires indépendantes, on a :

$$V(X + Y) = V(X) + V(Y)$$

Cette condition n'est pas suffisante pour traduire l'indépendance entre X et Y.

• Si X et Y sont deux variables aléatoires quelconques, on a :

$$V(X + Y) = V(X) + V(Y) + 2\,Cov(X ; Y)$$

E. Applications

Exemple 1

Exercice

Considérons l'expérience suivante : on lance deux dés équilibrés. A chaque couple de résultats, on associe le maximum X de l'une des composantes ou des deux composantes du couple.

1. Quelles sont les résultats possibles d'une telle expérience ?

2. Écrire la loi de probabilité de X et la représenter graphiquement.

3. Quelle est la fonction de répartition de X ? La représenter graphiquement.

Corrigé

1. Tous les résultats possibles (univers des possibles) d'une telle expérience figurent dans le tableau ci-dessous.

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

2. En associant, à chaque résultat, le **maximum** X du couple obtenu , on obtient les valeurs suivantes pour X :

$$X = \{1, 2, 3, 4, 5, 6\}$$

WWW.TALIB24.COM

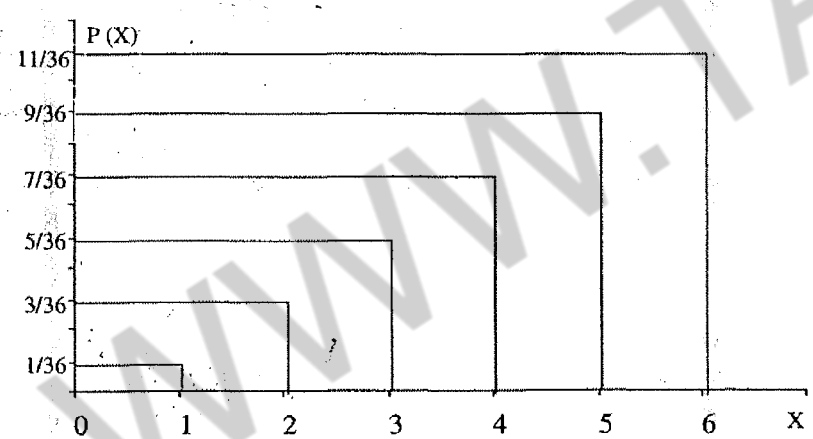
Pour chaque valeur de X, les probabilités correspondantes sont les suivantes (tableau ci-dessous).
Dans cet exemple, X ne peut prendre que des valeurs isolées ou entières : X est donc bien une variable aléatoire discrète.

Prob(X)	Couples correspondants
Prob(X = 1) = 1/36	(1,1)
Prob(X = 2) = 3/36	(2,1) ; (2,2) ; (1,2)
Prob(X = 3) = 5/36	(3,1) ; (3,2) ; (3,3) ; (2,3) ; (1,3)
Prob(X = 4) = 7/36	(4,1) ; (4,2) ; (4,4) ; (3,4) ; (2,4) ; (1;4)
Prob(X = 5) = 9/36	(5,1) ; (5,2) ; (5,3) ; (5;4) ; (5;5) ; (4,5) ; (3,5) ; (2,5) ; (1;5)
Prob(X = 6) = 11/36	ensemble grisé du tableau ci-dessus : 11 valeurs

• Écriture de la loi de probabilité de X

	x_i					
X	1	2	3	4	5	6
P(X)	1/36	3/36	5/36	7/36	9/36	11/36

• Représentation graphique de la loi de probabilité de X : diagramme en bâtons.

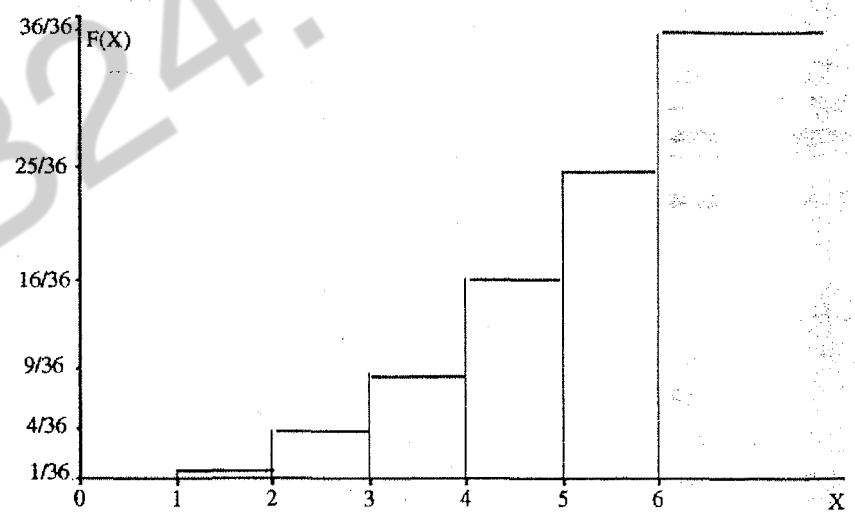


3. Fonction de répartition de X

• F : x ——> F(x) = Prob(X ≤ x)
autrement dit, c'est la probabilité que X prenne une valeur inférieure ou égale à une valeur donnée x.

	x								
X	- ∞	0	1	2	3	4	5	6	+ ∞
F (X)	0	0	0	1/36	4/36	9/36	16/36	25/36	36/36

• Représentation graphique de F(X) : diagramme en escalier



WWW.TALIB24.COM

Exemple 2

Dans l'exemple du § II.2, E (X) et V(X) seront déterminés ainsi :

• Espérance mathématique et variance

X	–2	–1	0	2	Total (= Σ)
Prob(X) ou p_i	3/36	12/36	15/36	6/36	1
X.Prob(X) ou $x_i p_i$	–6/36	–12/36	0	12/36	E (X) = –6/36
X^2. Prob(X) ou $x_i^2.p_i$	12/36	12/36	0	24/36	E (X^2) = 48/36

* $E(X) = \sum_{i=1}^{n} p_i x_i = -6/36 = -1/6$

* $V(X) = E(X^2) - [E(X)]^2 = 48/36 - 1/36 = 47/36$

3. Variable aléatoire continue

A. Fonction densité de probabilité : f(x)

Soit X, une variable aléatoire continue c'est-à-dire prenant ses valeurs dans R ou un intervalle de R.

On appelle **fonction de densité de probabilité de X** ou **densité de probabilité** de la variable aléatoire X, la **fonction f**, de R dans R^+ telle que, pour tout couple de nombres réels (a, b avec a ≤ b), la probabilité que X prenne des valeurs comprises entre a et b est :

$$\text{Prob}(a \le X \le b) = \int_a^b f(x)\,.\,dx$$

B. Fonction de répartition : F(x)

C'est l'application F : R ——> [0 ; 1] , tel que :

$$x \longrightarrow F(x) = \text{Prob}(X \le x) = \int_{-\infty}^{x} f(x).dx$$

Propriétés de f et F

Fonction de densité : f	Fonction de répartition : F
• f est intégrable • $\int_{-\infty}^{+\infty} f(x)dx = 1$ • f est positive	• F est dérivable • F ——> 0, quand x ——> – ∞ • F ——> 1, quand x ——> + ∞ • F est toujours croissante

C. Espérance mathématique

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$$

où f désigne la densité de X.

D. Variance

$$V(X) = \int_{-\infty}^{+\infty} (x-m)^2.f(x)dx$$

Les propriétés de E(X) et V(X) sont les mêmes que dans le cas d'une variable aléatoire discrète (cf. ci-dessus § I.3 et I.4)

WWW.TALIB24.COM

II. La loi binomiale

1. Position du problème

Soit une expérience dont le résultat est soit la réalisation d'un événement A soit celle de son contraire $\overline{A}$, toute autre éventualité étant évidemment exclue.
Associons à l'apparition de A, la probabilité p :

Prob(A) = p
Prob(A) + Prob($\overline{A}$) = 1 = p + q
Dans ce cas, Prob($\overline{A}$) = 1 – p = q

Exemple
Considérons un jeu consistant à jeter un dé à 6 faces et soient les événements suivants :
A = apparition de la face 3 ⇒ Prob (A) = 1/6 = p
$\overline{A}$ = apparition de « non A » c'est-à-dire {1,2,4,5,6} ⇒
Prob($\overline{A}$) = 1 – 1/6 = 5/6
Une expérience de ce type est appelée **épreuve de Bernouilli.**

Considérons maintenant une suite de **n épreuves** de Bernouilli **identiques et indépendantes.** A chaque épreuve, nous appellerons succès, la réalisation de A et échec la réalisation de $\overline{A}$.

Par exemple, si on jette le dé 5 fois, cela veut dire que **n = 5**

On voudrait calculer, sur ces 5 jets, la probabilité que la face 3 apparaisse :
– 1 fois sur les 5 épreuves
– 2 fois sur les 5 épreuves
– 3 fois sur les 5 épreuves
– 4 fois sur les 5 épreuves
– 5 fois sur les 5 épreuves
– ou alors qu'elle n'apparaisse aucune fois sur ces 5 épreuves. Ce qui est possible.

Appelons alors X, le nombre de succès, c'est-à-dire le nombre de fois où la face 3 est apparue lors de ces 5 jets.
X peut être considérée comme une **variable aléatoire** qui peut prendre les valeurs suivantes :

{0, 1, 2, 3, 4, 5}

Or, on sait que la probabilité associée à chaque apparition de la face 3 ne change pas : elle est toujours de 1/6.

Mais comment calculer la probabilité pour que la face 3 apparaisse 2 fois sur les 5 épreuves ?
La réponse est donnée par la loi binomiale !
Nous avons choisi cet exemple théorique uniquement pour mieux situer la position du problème mais on peut facilement deviner les situations concrètes où les problèmes se posent de la même façon (cf. exercices d'application).

2. Définition de la loi binomiale

Soient :
- A : événement aléatoire de probabilité p ; $\overline{A}$ → 1 – p [événement contraire]
- n : nombre d'épreuves où A est susceptible de se réaliser

WWW.TALIB24.COM

- X : variable aléatoire, associée à l'apparition de A, c'est-à-dire qu'on considère la v.a. X comme égale au nombre de fois où l'événement A a été réalisé au cours des n épreuves.

X peut prendre les valeurs [0, 1, 2, ... k, ...n]

La probabilité que A se produise k fois en n expériences (ou épreuves) est donnée par la loi binomiale :

$\text{Prob}(X = k) = C_n^k \, p^k q^{n-k}$ $\qquad [0 \leq k \leq n]$

ou,

$$\boxed{\textbf{Prob}(X = k) = \frac{n!}{k!\,(n-k)!} \, p^k q^{n-k}}$$

Cette loi de probabilité est appelée loi **binomiale** car $C_n^k \, p^k q^{n-k}$ est le terme général du développement du **binôme** de Newton $(p + q)^n$.

En effet, on a : $(p + q)^n = \sum_{k=0}^{n} C_n^k \, p^k \, q^{n-k}$

Cette loi est complètement définie par deux paramètres qui sont n et p. On la note alors $\mathcal{B}(n,p)$.

3. Propriétés de la loi binomiale

Moyenne : $E(X) = \mu = n.p$

Variance : $V(X) = \sigma^2 = n.p.q$

Écart type : $\sigma(X) = \sqrt{n.p.q}$

La démonstration de ces propriétés est évidemment possible mais ce qu'il faut retenir c'est surtout leurs **applications** qui sont très importantes.

Il existe des tables de la loi binomiale dont un extrait limité figure ci-dessous :

		p				
n	**k**	**0,1**	**0,2**	**0,3**	**0,4**	**0,5**
2	0	0,81	0,64	0,49	0,36	0,25
	1	**0,18**	0,32	0,42	0,48	0,5
	2	0,01	0,04	0,09	0,16	0,25
3	0	0,7290	0,512	0,343	0,216	0,125
	1	0,2430	0,384	0,441	0,432	0,375
	2	0,0270	0,096	0,189	0,288	0,375
	3	0,0010	0,008	0,027	0,064	0,125
4	0	0,6561	0,4096	0,2401	0,1296	0,0625
	1	0,2916	0,4096	0,4116	0,3456	0,2500
	2	0,0486	0,1536	0,2646	0,3456	0,3750
	3	0,0036	0,0256	0,0756	0,1536	0,2500
	4	0,0001	0,0016	0,0081	0,0256	0,0625

Exemple : pour n = 2 et k = 1 et p = 0,1
la table ci-dessus donne :

$\text{Prob}(X = 1) = C_n^k \, p^k q^{n-k} = 0{,}18$ (valeur grisée)

WWW.TALIB24.COM

4. Champ d'application de la loi binomiale

Dans la pratique, la loi binomiale, $\text{Prob}(X = k) = C_n^k\, p^k q^{n-k}$

est appliquée chaque fois où :

- un événement A se produit au cours d'**une** épreuve avec la probabilité p et l'événement contraire $\overline{A}$ avec la probabilité $\mathbf{q} = 1 - p$;
- l'événement A se répète **k fois** au cours de **n épreuves identiques** (à la première) et surtout **indépendantes :**
  - **épreuves identiques :** l'épreuve est toujours la même
  - **indépendantes :** le résultat d'une épreuve ne dépend pas du résultat de l'épreuve précédente ou de l'épreuve suivante.

Cette situation se présente dans les cas suivants :

- réussir **ou** non à un examen
- obtenir la face 2 lors d'un jet de dé **ou** non (c'est-à-dire obtenir 1, 3, 4, 5 ou 6)
- obtenir pile **ou** face lors d'un jet d'une pièce de monnaie
- une graine germe **ou** ne germe pas
- un animal est malade **ou** sain

Pour chacun des exemples cités, il existe toujours une probabilité p associée à l'apparition de l'événement considéré et une probabilité $q = 1 - p$, pour l'apparition de l'événement contraire.

Il est donc indispensable, lors de la résolution des exercices, de vérifier que l'on est bien dans la situation d'emploi de la loi binomiale en déterminant la nature des épreuves et les probabilités qui leur sont associées.

5. Application de la loi binomiale

Exemple

Exercice

D'une grande population d'animaux de la même race, dont 64% ne sont pas atteints d'une maladie A, on tire au hasard 16 d'entre eux *(on assimile ce tirage à un tirage avec remise)*. Soit X la variable aléatoire prenant pour valeur le nombre d'animaux n'ayant pas cette maladie :

1. Quelle est la loi de X ? Calculer son espérance E(X) et sa variance V(X).

2. **a.** Quelle est la probabilité d'avoir 10 d'entre eux non atteints ?

b. Quelle est la probabilité d'avoir au plus 15 d'entre eux non atteints ?

Corrigé

1. Le tirage est effectué 16 fois dans des conditions identiques et indépendantes, chaque tirage donne lieu a deux résultats possibles qui s'excluent mutuellement, (ou ils sont atteints de la maladie A ou ils ne sont pas atteints). La variable aléatoire X est donc une v.a binomiale de paramètres n = 16 et p = 0,64

Sa loi de probabilité est donc : B(16 ; 0,64)

$$P(X = k) = C_{16}^k (0,64)^k (0,36)^{16-k}$$

$0 \leq X \leq 16$

$E(X) = np = 16 \times 0,64 = 10,24$

$V(X) = npq = 16 \times 0,64 \times 0,36 = 3,6864$

2.

a. $P(X = 10) = C_{16}^{10}(0,64)^{10} \times (0,36)^6 = 0,2010$

b. P*(au plus 15 non atteints)* = 1 – P*(16 non atteints)*

$= 1 - C_{16}^{16}(0,64)^{16}(0,36)^0$

$= 1 - (0,64)^{16} = 1 - 0,0008 = 0,9992$

WWW.TALIB24.COM

III. La loi de poisson

1. Définition de la loi de Poisson

Considérons une v.a. X suivant une loi binomiale $\mathcal{B}$ (n, p) **mais où n est très grand et p voisin de 0.**

Exemple

Considérons la loi $\mathcal{B}$ (100, 0,02)

$$\text{Prob}(X = k) = C_{100}^{k}\, 0{,}02^{k} \cdot 0{,}98^{100-k} \qquad [0 \leq X \leq 100]$$

Les probabilités correspondant aux valeurs élevées de X sont extrêmement faibles et pratiquement négligeables. Les valeurs significatives sont groupées autour de :

$$k = np = 100 \times 0{,}02 = 2$$

On remarque que, la valeur de n étant **élevée** et celle de p étant **faible**, l'expression binomiale devient peu pratique.

On montre alors qu'une très bonne approximation des diverses probabilités peut être obtenue au moyen de la **loi de Poisson** de paramètre m = np = 2 soit :

$$\text{Prob}\,(X = k) \approx e^{-2}\,\frac{2^k}{k\,!}$$

L'approximation que l'on fait correspond en fait à une propriété de la loi binomiale.

En effet, on montre que :

Quand $n \longrightarrow \infty$, et $p \longrightarrow 0$ et $np = m$ restant constant, alors :

$$C_n^k\; p^k q^{n-k} \longrightarrow e^{-m}\frac{m^k}{k\,!}$$

En pratique ceci est vérifié lorsque **n > 30 et p < 0,1** ou **np < 5** ou **npq < 10.** Dans ce cas l'approximation ci-dessus est acceptable.

Règle générale

Soit X une variable aléatoire pouvant prendre toutes les valeurs entières 0,1, 2, ...n,...

On dit que X suit la loi de Poisson de paramètre m (m réel strictement positif) si et seulement si :

$$\boxed{\text{Prob}\,(X = k) = e^{-m} \times \frac{m^k}{k\,!}}$$

avec k = 0, 1, 2,.....n,...

et où **e** est la base des logarithmes népériens [e = 2,718...].

Cette loi est définie par un seul paramètre, **m.** On la note **P (m).**

2. Propriétés de la loi de Poisson et utilisation des tables

$$E(X) = m = np$$
$$V(X) = \sigma^2 = m$$
$$\sigma\,(X) = \sqrt{m}$$

A. Quand utilise-t-on la loi de Poisson ?

La loi de Poisson est utilisée lorsque la réalisation d'un événement est très rare sur un grand nombre d'observations. C'est le cas de nombreux phénomènes tels que :

- nombre d'appels téléphoniques reçus par minute par les services de renseignements téléphoniques ;
- nombre de fautes d'impression par page dans un livre ;
- désintégration d' un corps radioactif ;
- présence de globules rouges dans un champ microscopique après dilution ;

WWW.TALIB24.COM

– présence d'une bactérie dans un champ microscopique après dilution d'une suspension bactérienne ;
– etc.

B. Utilisation des tables

Certains ouvrages de statistique donnent des tables de la loi de Poisson. Ces tables correspondent soit à la fonction de distribution Prob(X = k) (table 1 en fin de ce chapitre), soit à la fonction de répartition Prob(X ≤ k) (table 2 en fin de ce chapitre).
Ces tables sont données en fonction de l'espérance mathématique, m et de k.

3. Approximation d'une loi binomiale par une loi de Poisson

A. Conditions préalables

La loi de Poisson est un cas particulier d'une loi binomiale où n est très grand (n > 50) et la probabilité p très petite (p < 0,1), le produit np conservant une valeur finie m.
Donc, si X est une variable aléatoire dont la loi de probabilité est une loi binomiale $\text{Prob}(X = k) = C_n^k\, p^k q^{n-k}$ et si n augmente indéfiniment de telle sorte que le produit np reste égal à une constante m ≠ 0 alors,

$$\text{Prob}(X = k) \text{ tend vers } e^{-m}\frac{m^k}{k!}$$

Cette approximation simplifie considérablement les calculs lorsqu'elle est possible.
Elle ne s'applique cependant que si n > 30 (ou 50) et si p < 0,1.

B. Exemple

Exercice

Une usine fabrique des articles dont 4 % présentent des défauts. Quelle est la probabilité pour que dans une livraison de 75 de ces articles, il y ait 2 articles défectueux ?

Corrigé

La probabilité élémentaire pour qu'un article pris au hasard dans cette usine soit défectueux est p = 0,04.
Autrement dit, la probabilité d'avoir 2 articles défectueux dans un lot de 75 est donnée par la loi binomiale de paramètres :

$$n = 75 \quad \text{et} \quad p = 0{,}04$$

donc, pour x = 2, on a :

$$\text{Prob}(X = 2) = C_{75}^2 (0{,}04)^2 (0{,}96)^{75-2}$$

Nous sommes dans un cas où la loi binomiale peut paraître difficile à appliquer ; comme n est « grand » et m = n.p < 10, on peut approximer cette loi binomiale $\mathcal{B}(75 ; 0{,}04)$ par une loi de Poisson, P(m) = P(n.p) = P(3)
La probabilité cherchée sera donc, pour m = 3 et k = 2 :

$$P(X = 2) = \frac{3^2}{2!} e^{-3} = 0{,}224 \quad \text{(table 1, en fin de chapitre).}$$

WWW.TALIB24.COM

Table 1. Loi de Poisson. Fonction de distribution :

$$\text{Prob}(X = k) = e^{-m} \cdot \frac{m^k}{k\,!}$$

m / k	0,1	0,3	0,5	0,7	1	2	3	4	5	8	10
0	0,9048	0,7408	0,6065	0,4966	0,3679	0,1353	0,0498	0,0183	0,0067	0,0003	0,0000
1	0,0905	0,2222	0,3033	0,3476	0,3679	0,2707	0,1494	0,0733	0,0337	0,0027	0,0005
2	0,0045	0,0333	0,0758	0,1217	0,1839	0,2707	0,2240	0,1465	0,0842	0,0107	0,0023
3	0,0002	0,0033	0,0126	0,0284	0,0613	0,1804	0,2240	0,1954	0,1404	0,0286	0,0076
4		0,0003	0,0016	0,0050	0,0153	0,0902	0,1680	0,1954	0,1755	0,0573	0,0189
5			0,0002	0,0007	0,0031	0,0361	0,1008	0,1563	0,1755	0,0916	0,0378
6				0,0001	0,0005	0,0120	0,0504	0,1042	0,1462	0,1221	0,0631
7					0,0001	0,0034	0,0216	0,0595	0,1044	0,1396	0,0901
8						0,0009	0,0081	0,0298	0,0653	0,1396	0,1126
9						0,0002	0,0027	0,0132	0,0363	0,1241	0,1251
10							0,0008	0,0053	0,0181	0,0993	0,1251
11							0,0002	0,0019	0,0082	0,0722	0,1137
12							0,0001	0,0006	0,0034	0,0481	0,0948
13								0,0002	0,0013	0,0296	0,0729
14								0,0001	0,0005	0,0169	0,0521
15									0,0002	0,0090	0,0347
16										0,0045	0,0217
17										0,0021	0,0128
18										0,0009	0,0071
19										0,0004	0,0037
20										0,0002	0,0019
21										0,0001	0,0009
22											0,0004
23											0,0002

Table 2. Loi de Poisson. Fonction de répartition :

$$\text{Prob}(X \le k) = \int_{-\infty}^{k} e^{-m} \cdot \frac{m^k}{k\,!}\, dk$$

m / k	0,1	0,3	0,5	0,7	1	2	3	4	5	8	10
0	0,9048	0,7408	0,6065	0,4966	0,3679	0,1353	0,0498	0,0183	0,0067	0,0003	0,0000
1	0,9953	0,9631	0,9098	0,8442	0,7358	0,4060	0,1991	0,0916	0,0404	0,0030	0,0005
2	0,9998	0,9964	0,9856	0,9659	0,9197	0,6767	0,4232	0,2381	0,1247	0,0138	0,0028
3	1,0000	0,9997	0,9982	0,9942	0,9810	0,8571	0,6472	0,4335	0,2650	0,0424	0,0103
4		1,0000	0,9998	0,9992	0,9963	0,9473	0,8153	0,6288	0,4405	0,0996	0,0293
5			1,0000	0,9999	0,9994	0,9834	0,9161	0,7851	0,6160	0,1912	0,0671
6				1,0000	0,9999	0,9955	0,9665	0,8893	0,7622	0,3134	0,1301
7					1,0000	0,9989	0,9881	0,9489	0,8666	0,4530	0,2202
8						0,9998	0,9962	0,9786	0,9319	0,5925	0,3328
9						1,0000	0,9989	0,9919	0,9682	0,7166	0,4579
10							0,9997	0,9972	0,9863	0,8159	0,5830
11							0,9999	0,9991	0,9945	0,8881	0,6968
12							1,0000	0,9997	0,9980	0,9362	0,7916
13								0,9999	0,9993	0,9658	0,8645
14								1,0000	0,9998	0,9827	0,9165
15									0,9999	0,9918	0,9513
16									1,0000	0,9963	0,9730
17										0,9984	0,9857
18										0,9993	0,9928
19										0,9997	0,9965
20										0,9999	0,9984
21										1,0000	0,9993
22											0,9997
23											0,9999

WWW.TALIB24.COM

Testez vos connaissances

Exercice 1

Un tissu biologique est contaminé par un virus et l'observation porte sur 10000 lots de 30 fragments.

1. Exprimer les probabilités théoriques de trouver 20 fragments sur 30, puis 21 fragments contaminés sur 30. On appelle p la probabilité qu'un fragment soit contaminé et q la probabilité de l'événement contraire.

2. On remarque que sur 30 fragments, on trouve:
- 20 fragments contaminés dans 357 lots sur 10 000
- 21 fragments contaminés dans 210 lots sur 10 000

En déduire les valeurs numériques de p et q.

Exercice 2

Le groupe sanguin d'un individu peut être AB, B, O ou A. Sur 400 personnes, 20 possèdent le groupe AB, 40 le groupe B, 160 le groupe O et 180 le groupe A. D'autre part, chaque individu peut avoir un rhésus + (85 % des personnes) ou un rhésus – (15 %). La possibilité de recevoir du sang sans accident dépend de deux caractères indépendants: le groupe sanguin et le facteur rhésus. En ce qui concerne le groupe sanguin, on utilisera le tableau en page suivante.

WWW.TALIB24.COM

Groupe	Peut donner du sang sans accident à :	Peut recevoir du sang sans accident de :
AB	AB	tous les groupes
A	AB et A	O et A
B	AB et B	O et B
O	tous les groupes	O

En ce qui concerne le facteur rhésus, on admettra qu'une personne de rhésus + peut recevoir sans accident du sang d'un donneur avec rhésus + ou – , et qu'une personne de rhésus – ne peut recevoir sans accident du sang que d'un donneur avec rhésus – .

Déterminer :

1. La probabilité Prob(B) pour qu'un individu pris au hasard possède le groupe B.
La probabilité Prob(B–) pour qu'il possède le groupe B et le rhésus - .

2. La probabilité Prob(O)
La probabilité Prob(O+)

3. La probabilité pour un individu B– de recevoir sans accident du sang d'un autre individu pris au hasard.
Même question pour un individu O +

4. On réunit trois personnes au hasard et on prélève sur elles du sang pour donner à un individu du groupe B– .
Quelle est la probabilité que pour deux d'entre elles, on soit sûr que l'opération se fasse sans accident ?

Exercice 3

Dans un élevage de bovins viande, on a constaté qu'en moyenne 10 animaux sur 1000 sont malades et donc non commercialisables.
Sur un lot de 800 animaux :
1. Calculer la probabilité d'avoir 4 animaux malades.
2. Calculer la probabilité d'avoir moins de 3 animaux malades.

Exercice 4

Dans une certaine population de bovins lait, 6 % d'entre eux sont atteints d'une mammite. On prélève un échantillon de 50 individus : X désigne le nombre d'animaux malades dans cet échantillon.
1. Peut-on utiliser l'approximation de Poisson dans ce cas ?
2. Dresser un tableau des valeurs que peut prendre X.
3. Calculer la probabilité pour que 3 animaux au plus soient malades.

Exercice 5

On considère N portées de cinq souriceaux, issues de couples hétérozygotes. On rappelle que la coloration des souris peut s'expliquer par l'existence d'un gène de coloration susceptible de 2 états G et b, G entraînant la coloration, b l'albinisme. Pour un individu, il y a donc 3 génotypes possibles :
– individu homozygote coloré GG,
– individu homozygote albinos bb,
– individu hétérozygote coloré Gb.

WWW.TALIB24.COM

Le croisement de deux individus hétérozygotes Gb peut donner lieu à l'une ou l'autre des 4 combinaisons possibles GG, Gb, bG ou bb, avec des probabilités respectives égales à 1/4.
Soit X, la variable aléatoire : nombre de portées comportant au moins 4 souriceaux gris.

1. Quelle loi suit X ? La justifier.
2. En déduire la moyenne, la variance et l'écart type de X pour n = 4
3. Calculer les probabilités pour que X = 0,1,2,3,4

Vérifiez vos réponses à la page suivante

Réponses au test

Exercice 1

1. Appliquons la loi binomiale
$\text{Prob}(20) = C_{30}^{20} (p)^{20} (q)^{10}$
$\text{Prob}(21) = C_{30}^{21} (p)^{21} (q)^{9}$

2. Sachant que,

$$\text{Prob}(20) = \frac{357}{10\,000} = 0{,}0357$$

et

$$\text{Prob}(21) = \frac{210}{10\,000} = 0{,}0210$$

la valeur numérique de la probabilité élémentaire est déterminée à partir du rapport

$$\frac{\text{Prob}(20)}{\text{Prob}(21)} = \frac{357}{210} = \frac{C_{30}^{20}(p)^{20}(q)^{10}}{C_{30}^{21}(p)^{21}(q)^{9}} = 2{,}1 \times \frac{q}{p} = 2{,}1\ \frac{1-p}{p}$$

d'où $p = 0{,}55$ et $q = 0{,}45$

Exercice 2

1. Prob(B) = 40/400 = 0,1
Prob (B–) = 0,1 × 0,15 = 0,015

2. Prob(O) = 160/400 = 0,4
Prob(O+) = 0,4 × 0,85 = 0,34

WWW.TALIB24.COM

3. Un individu B– peut recevoir sans accident du sang d'un sujet O– ou B– ; d'où la probabilité cherchée :

Prob(O–) + Prob(B–) = 0,5 × 0,15 = 0,075

Un individu O+ peut recevoir sans accident du sang d'un sujet O quel que soit son rhésus ;

d'où la probabilité cherchée : 0,4

4. La probabilité pour un individu B– de recevoir du sang sans accident est : p = 0,075

Appliquons la loi binomiale :

$$\text{Prob}(x) = C_n^x (p)^x (q)^{n-x}$$

avec x = 2, n = 3, p = 0,075 et q = 0,925

on a, p(2) = 0,0156 = 1,56 %

Exercice 3

On peut considérer le nombre d'animaux malades comme étant une v.a. X qui suit une loi binomiale de paramètres **n = 800** et **p = 0,01** (c'est-à-dire 10/1000).

Comme n est grand et p est petit, on **peut donc approximer la loi B (800, 0,01) par la loi de Poisson P (m)** avec m = n.p = 800.0,01 = 8

1. $\text{Prob}(X = k) = e^{-m} \cdot \frac{m^k}{k\,!} = e^{-8} \cdot \frac{8^4}{4\,!} = \mathbf{0{,}057}$ (table 1)

2. Pour avoir moins de 3 animaux malades, il faut avoir 0, 1 ou 2 animaux malades c'est-à-dire :

$$P_0 + P_1 + P_2 = e^{-8} \cdot \frac{8^0}{0\,!} + e^{-8} \cdot \frac{8^1}{1\,!} + e^{-8} \cdot \frac{8^2}{2!}$$

$= e^{-8}\,[1 + 8 + 64/2] = \mathbf{0{,}0138}$ (tables 1 et 2 pages 62 et 63 pour x ≤ 2)

Exercice 4

1. p = 0,06 et n = 50 ⇒ n.p = 0,06 × 50 = 3

On peut considérer X comme une variable aléatoire qui suit une loi binomiale de paramètres n = 50 et p = 0,06.

Comme p est voisin de 0 et np < 5, on peut assimiler B(n, p) à une loi de Poisson de paramètre m = n.p = 3

2. Tableau des valeurs de X : X variant de 0 à 50

Nous avons : $\text{Prob}(X = k) = e^{-m} \cdot \frac{m^k}{k\,!} = e^{-3} \cdot \frac{3^k}{k\,!}$

Au lieu de calculer les différentes valeurs des probabilités pour k = 0, 1, 2, 3… à l'aide de cette formule, il suffit de lire ces valeurs dans la table 1 ci-dessus en prenant la colonne **m = 3** et les lignes correspondant aux différentes valeurs de k.

Exemple :

Prob(k = 0) = 0,0498

Prob(k = 1) = 0,1494

etc.

3. Calculer la probabilité pour que 3 animaux au plus soient malades revient à calculer

Prob (X ≤ 3) = Prob(X = 0) + Prob(X=1) + Prob(X=2) + Prob(X=3

= 0,0498 + 0,1494 + 0,224 + 0,224 = **0,6472**

(cf. fin du présent chapitre, table 1)

Mais on peut aussi trouver cette valeur directement dans la table 2 (cf. fin du présent chapitre) qui représente les valeurs cumulées.

Prob (≤ 3) = 0,6472

WWW.TALIB24.COM

Exercice 5

1. La coloration d'une souris s'expliquant par le croisement suivant :

	G	b
G	GG	Gb
b	Gb	bb

Les souriceaux de gènes GG,Gb, bG sont colorés, les bb sont albinos.

D'où la probabilité pour un souriceau d'être coloré = 3/4 (tableau)

$$p = 3/4 \qquad \Rightarrow \qquad q = 1/4$$

La probabilité d'avoir au moins 4 souriceaux gris dans une portée de 5 souriceaux est :

$$\text{Prob}(4) = P = C_5^4 (3/4)^4 (1/4)^1 + C_5^5 (3/4)^5 (1/4)^0$$

d'où : $\text{Prob} = 0{,}395 + 0{,}237 = 0{,}632$

$$Q = 1 - P = 1 - 0{,}632 = 0{,}368$$

La variable aléatoire X suit la loi binomiale :

$$\text{Prob}(X) = C_N^X (0{,}632)^X (0{,}368)^{N-X}$$

2. $m = NP = 0{,}632 \times 4 = 2{,}528$

$V = NPQ = 2{,}528 \times 0{,}368 = 0{,}930$

$s = \sqrt{0{,}930} = 0{,}9645$

3. Calculons les probabilités pour X = 0, 1, 2, 3, 4 respectivement

$\text{Prob}(0) = C_4^0 (0{,}632)^0 (0{,}368)^4 = 0{,}018$

de même,

$\text{Prob}(1) = 0{,}126$

$\text{Prob}(2) = 0{,}326$

$\text{Prob } 3) = 0{,}369$

$\text{Prob}(4) = 0{,}159$

WWW.TALIB24.COM

Exercices supplémentaires

Exercice 1

Le pourcentage des succès à un examen est de 40 %. Quelle est la probabilité pour que, dans un groupe de 5 candidats,

1. aucun ne soit admis
2. un et un seul soit admis
3. 2 et seulement 2 soient admis
4. au moins 2 soient admis
5. les 5 soient admis

Exercice 2

Une usine fabrique des bouteilles d'huile dont 2 % sont défectueuses. Cette usine livre 10 000 bouteilles à un client.
Soit X, la variable aléatoire « nombre de bouteilles défectueuses dans cette livraison ».

1. Quelle est la loi de probabilité de X ? Quels sont ses paramètres ?
2. Calculer la moyenne la variance et l'écart type de X en interprétant vos résultats.

Exercice 3

Sachant que la probabilité d'avoir un garçon est p = 0,48 ; quelle est, dans une famille de 5 enfants, la probabilité d'avoir 2 garçons ?

WWW.TALIB24.COM

Exercice 4

On pose à un candidat une question en lui proposant 3 réponses (dont la réponse correcte). Parfaitement ignorant, ce candidat donne au hasard l'une des 3 réponses proposées.

1. Quelle est la probabilité pour qu'il donne la réponse correcte ?

2. Un test d'aptitude se compose de 5 questions posées dans les conditions ci-dessus.
Quelle est la probabilité pour que notre candidat donne les réponses à 3 questions ; à 4 questions ; à 5 questions ?

3. Avec quelle probabilité notre candidat sera-t-il reconnu apte, s'il faut, pour cela, qu'il donne les réponses correctes à 3 questions au moins ?

4. Avec quelle probabilité notre candidat sera-t-il reconnu apte, s'il connaît les réponses correctes à 2 des 5 questions ?

Exercice 5

Une machine à embouteiller peut tomber en panne. La probabilité d'une panne est de 0,01 à chaque emploi de la machine. La machine doit être utilisée 100 fois.
Le nombre de pannes obtenues est une variable aléatoire X.

1. Énoncer sa loi de probabilité exacte et calculer les probabilités d'obtenir 0, 1, 3 pannes ou 4 pannes et plus.

2. Énoncer une loi approchée de la variable aléatoire X et calculer avec l'approximation des tables, les probabilités d'obtenir les mêmes conditions sur la valeur de X.

3. On estime le coût d'une réparation à 500 F. La dépense, exprimée en F, pour réparation de la machine est une variable aléatoire Y. Calculer l'espérance mathématique de Y et sa variance. On assimilera le coût de 4 pannes ou plus au coût de 4 pannes exactement.

WWW.TALIB24.COM

Corrigé des exercices supplémentaires

Exercice 1

$n = 5$
$p = 0{,}4$ d'où $q = 1 - p = 0{,}6$
Prob (k succès) = $C_5^k\,(0{,}4)^4\,(0{,}6)^{5-k}$

1. Prob(k = 0) = 7,8 %
2. Prob(k = 1) = 25,9 %
3. Prob(k = 2) = 34,6 %
4. $2 \leq k \leq 5$ Prob = 1 – Prob(0) – Prob(1) = 66,3 %
5. Prob (k = 5) = 1 %

Exercice 2

1.
$p = 0{,}02$, donc $q = 1 - p = 1 - 0{,}02 = 0{,}98$

$n = 10\,000$ et $p = 0{,}02$

$$\text{Prob}(X = k) = C_n^k\, p^k q^{n-k}$$

2.

• Calcul de la moyenne m
$m = n.p = 10\,000 \times 0{,}02 = 200$ bouteilles
• Calcul de la variance V(X)
$\sigma^2 = n.p.q = 10\,000 \times 0{,}02 \times 0{,}98 = 196$

• Calcul de l'écart type
$\sigma = \sqrt{\sigma^2} = \sqrt{196} = 14$ bouteilles
• Interprétation des résultats
Si la fabrique réalise, dans les conditions de l'énoncé, des livraisons de 10 000 bouteilles, il y a en moyenne 200 d'entre elles qui sont défectueuses pour une livraison.
De plus les dispersions de bouteilles défectueuses par rapport à la moyenne seraient de ± 14, autrement dit l'écart que subirait X serait :
au mieux X = 200 – 14 = 186 bouteilles
au pire X = 200 + 14 = 214 bouteilles.

Exercice 3

$N = 5$
$x = 2$
$p = 0{,}48 \Rightarrow q = 0{,}52 \Rightarrow \text{Prob}(x=2) = C_5^2\,(0{,}48)^2\,(0{,}52)^3$
$= 10.0{,}2304.0{,}1406 = \mathbf{0{,}324}$

Exercice 4

1. Calcul de la probabilité, p, de donner la réponse correcte. On a évidemment : **p = 1/3**

2. Calcul de Prob(3), Prob(4) et Prob(5). On a, d'après la loi binomiale
$\text{Prob}(k) = C_5^k\,(1/3)^k\,(2/3)^{5-k}$
Prob(3) = 16,5 %
Prob(4) = 4,1 %
Prob(5) = 0,4 %

WWW.TALIB24.COM

3. Calcul de la probabilité P, d'être reconnu apte.

Prob = Prob(3) + Prob(4) + Prob(5) soit **Prob = 21 %**

4. Étude du cas où le candidat connaît 2 questions sur 5.

Pour être reconnu apte, le candidat doit donner au moins une réponse exacte sur les 3 questions qu'il ne connaît pas. La probabilité cherchée est donc $P' = 1 - (2/3)^3$ car $(2/3)^3$ exprime la probabilité de ne donner aucune réponse exacte aux questions non étudiées.

Donc, **Prob'** = 1 – 8/27 = 19/27 = **70,4 %**

Exercice 5

1. On peut alors considérer X comme une v.a. qui suit une **loi binomiale** de paramètres **n = 100** (nombre de fois où la machine est employée) et **p = 0,01** (probabilité de panne) donc : **q = 1 – p = 0,99**

Dans ces conditions, on a :

Prob(X = 0) = 0,3660

Prob(X = 1) = 0,3697

Prob(X = 2) = 0,1849

Prob(X = 3) = 0,0610

Prob(X ≥ 4) = 1 – [Prob(X=0) + Prob(X=1) + Prob(X=2) + Prob(X=3)]

= 1 – [0,3660 + 0,3697 + 0,1849 + 0,0610]

= 1 – 0,9816 = 0,0184

2. p ≤ 0,1 et n > 50. On peut donc approximer la loi binomiale précédente par la loi de Poisson de paramètre m = np = 100 × 0,1 = 1.

On a alors, voir table 1 de Poisson

Prob(X = 0) = 0,3679

Prob(X = 1) = 0,3679

Prob(X = 2) = 0,1839

Prob(X = 3) = 0,0613

et d'après la table 2, fonction de répartition de Poisson et la table 1, fonction de distribution de Poisson :

Prob(X ≥ 4) = [1 – Prob(X ≤ 4)] + Prob(X= 4) = [1 –0,9963] + 0,0153 = 0,019

3. A chaque valeur de X, on associe une valeur de Y et une probabilité :

X	0	1	2	3	4
Y	0	500	1000	1500	2000
p	0,3679	0,3679	0,1839	0,0613	0,019

$$\mathbf{E(Y)} = \sum_{i=1}^{5} p_i . y_i$$

= 0,3679 × 0 + 0,3679 × 500 + 0,1839 × 1000 + 0,0613 × 1500 + 0,019 × 2000 = **497,50**

$$\mathbf{V(Y)} = E(Y)^2 - [E(Y)]^2$$

$E(Y)^2 = 0{,}3679 \times 0^2 + \ldots + 0{,}019 \times 2000^2 = 489\ 250$

$V(Y) = 489\ 250 - 497{,}5^2 =$ **241 744**

WWW.TALIB24.COM

SÉQUENCE DE TRAVAIL N°3

La loi normale ou loi de Gauss $\mathcal{N}(m, \sigma)$

3

Objectifs pédagogiques

A la fin de cette séquence, mais étape par étape, vous devriez être capable :

1. de définir et de décrire la distribution d'une loi normale ;
2. d'utiliser les tables relatives à la loi normale ;
3. de définir l'intervalle de confiance d'une moyenne ;
4. d'établir la relation entre la loi binomiale et la loi normale ;
5. d'utiliser la loi normale pour résoudre des problèmes concrets.

4 h

Attention ! Cette séquence est un prérequis indispensable à la suite du cours. Vous ne pouvez aborder les séquences suivantes que si vous avez bien assimilé cette partie.

WWW.TALIB24.COM

De quoi s'agit-il ?

Pour mieux comprendre cette partie du cours, commencez par bien lire cette page...

L'appréciation du squelette d'une vache normande, en vue de son indexation, s'effectue à partir de 4 postes du squelette et 3 notes de synthèse pour les trois ensembles anatomiques suivants : squelette, musculature et mamelle.

La taille, par exemple, est notée de 1 à 9 en fonction de la hauteur au sacrum (HS) et de l'âge de l'animal. A chaque point supplémentaire correspond un gain de 2 cm au niveau de la mensuration, la moyenne de la race représentant la note intermédiaire 5.

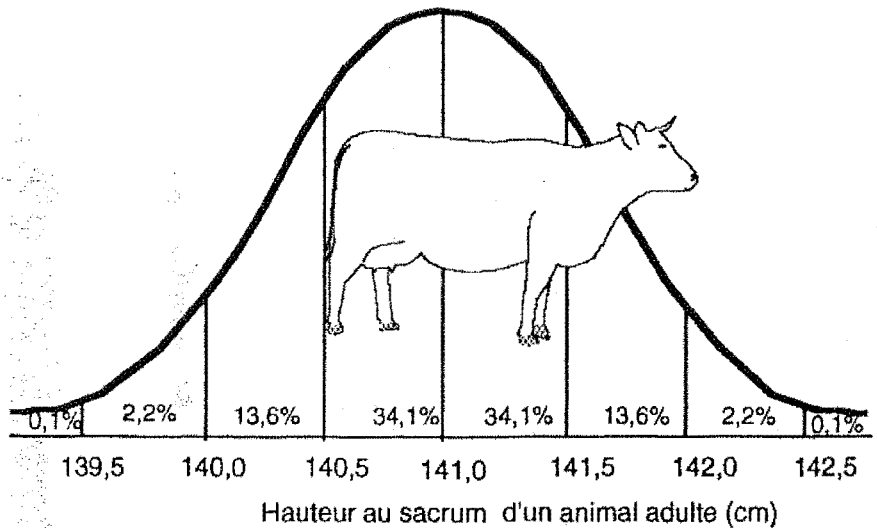
Moyenne de la race			**Objectif à l'âge adulte**
à 2 ans	à 3 ans	adulte	
138 cm	140 cm	142 cm	147 cm

La hauteur au sacrum est mesurée à la toise au niveau des vertèbres caudales, derrière la croisée des hanches, l'animal se tenant de profil et bien posé sur ses aplombs. Elle remplace la hauteur au garrot jugée moins « répétable » car plus dépendante de l'attitude de l'animal.

Petite à très petite	Moyenne	Grande à très grande
HS ≤ 138 cm	139 ≤ HS ≤ 143 cm	HS ≥ 144 cm

WWW.TALIB24.COM

Une enquête (exemple fictif) faite en Normandie, a permis d'établir la courbe « en cloche » ci-dessous.



I. Étude préliminaire

Que remarque-t-on sur la courbe ?

Cette courbe représente la relation entre les résultats obtenus aux mensurations de la hauteur au sacrum (HS) et les classes de population qui leur sont associés. Ainsi, **l'échantillon** sur lequel a été effectuée l'étude - assez **grand** pour être considéré comme une **population** - montre que :

- la majorité de la population a une HS qui varie entre 140,5 et 141,5 cm ; cette majorité représente :

  34,1 % (140,5 cm ≤ HS ≤ 141,0 cm) + **34,1 %** (141,0 cm ≤ HS ≤ 141,5 cm) = **68,2 %**

- les individus dont la HS est inférieure à 139,5 cm (les plus " petits ") ne représentent plus que **0,1 %** de la population ;
- les individus dont la HS est supérieure à 142,5 cm (les plus " grands ") ne représentent également que **0,1 %** de la population.

La grande majorité de la population se concentre donc autour de la moyenne (HS = 141,0 cm). Plus on s'éloigne de cette moyenne et plus les individus sont rares.

On dit que la distribution des résultats obtenus par cette mensuration obéit une loi normale représentée par la courbe en cloche, appelée aussi courbe de GAUSS ou courbe de la loi normale.

Lorsque la valeur que prend un caractère a des causes très nombreuses et indépendantes les unes des autres, les valeurs de ce caractère peuvent être représentées par cette courbe. C'est le cas de la hauteur au sacrum dont on pense qu'il y a de multiples raisons qui « expliquent » le résultat obtenu : facteurs génétiques, environnement...

Si vous êtes intéressés par le concept de la loi normale, nous vous conseillons de lire l'article de la revue POUR LA SCIENCE en annexe du livret.

WWW.TALIB24.COM

II. Distribution de Gauss ou loi normale : $\mathcal{N}(m, \sigma)$

1. Définition

Soit X une variable aléatoire **continue**, prenant ses valeurs dans $\mathbb{R}$, de moyenne m et d'écart type = σ.
On dit que X obéit à une loi normale ou loi de Gauss, si sa **densité de probabilité** est :

$$f(X = x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-m}{s}\right]^2} \qquad -\infty < x < +\infty$$

x étant une valeur numérique particulière de X,
et e = 2,718 est la base des logarithmes népériens.

La **fonction de répartition** de X est :

$$F(X) = \text{Prob}(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$$

ou,

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left[\frac{x-m}{s}\right]^2} dx$$

La somme des probabilités relatives à toutes les valeurs de x est égale à 1, c'est-à-dire :

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left[\frac{x-m}{\sigma}\right]^2} dx = 1$$

La fonction de répartition dépend de **deux paramètres m et σ**. Elle est notée habituellement par $\mathcal{N}(m, \sigma)$.

2. Loi Centrée réduite $\mathcal{N}(0, 1)$

La fonction précédente étant peu maniable pratiquement, on peut simplifier son utilisation en procédant à un changement de variable.
Pour cela, définissons une **nouvelle variable u** telle que :

$$u = \frac{x-m}{\sigma}$$

On définit alors une **nouvelle fonction de densité de probabilité** de cette nouvelle variable :

$$f(u) = Y = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-1/2u^2}$$

et sa **fonction de répartition** devient :

$$\Phi(u) = \text{Prob}(U \leq u) = \int_{-\infty}^{u} f(u)\, du$$

ou,

$$\Phi(u) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{u} e^{-1/2u^2} du$$

Comme **m = 0** et σ = 1, la variable est dite **centrée** (c'est-à-dire de moyenne nulle) et **réduite** (c'est-à-dire son écart type vaut 1). Elle est notée $\mathcal{N}(0, 1)$.

> En fait, ce changement de variable correspond tout simplement à un changement d'échelle et à une translation sur l'axe des abscisses. On obtient ainsi une densité f(u) indépendante de m et s ; l'intérêt de ce changement réside alors dans le fait qu'on puisse utiliser la même table pour des variables aléatoires obéissant à des lois normales de différents paramètres.

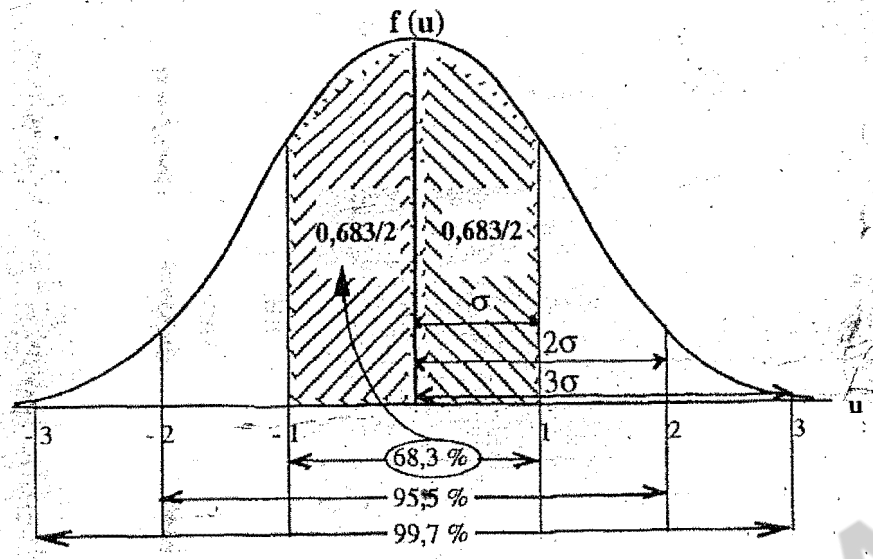
La représentation graphique de la fonction densité de probabilité **f(u)**, est une courbe en « cloche de Gauss » f(u) tend rapidement vers 0 quand u augmente (courbe page suivante).

WWW.TALIB24.COM

On montre que :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \; e^{-1/2\, u^2} \, du = 1$$

Nous verrons très bientôt l'importance de cette propriété.



Voici quelques valeurs remarquables de la courbe

u	–3 (ou + 3)	–2 (ou +2)	–1 (ou + 1)	0	0,5 (ou –0.5)
f(u)	0,004	0,054	0,242	0,399	0,352

Il existe des tables aussi bien pour la fonction densité de probabilité f(u) que pour la fonction de répartition Φ(u).

• **La table de la fonction densité de probabilité**, dont un extrait est présenté ci-dessous, donne, pour chaque valeur de u (addition des nombres inscrits en marge), la valeur :

$$f(u) = \frac{1}{\sqrt{2\pi}} \; e^{-1/2\, u^2}$$

Extrait de la table de la distribution normale réduite :
fonction densité de probabilité f (u).

Exemple : si u = 0,55 = (0,5 + 0,05) alors f(u) = 0,3429

u	0,00	0,01000	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315

WWW.TALEBEC.COM

• **La table de la fonction de répartition**, donne les aires comprises sous la courbe normale entre u = - ∞ et toute valeur positive de u : c'est la **table I**, de la fonction de répartition (cf. à la fin du livret) dont un extrait est présenté ci-dessous. A partir de cette table on peut obtenir l'aire comprise entre 2 valeurs quelconques de u, en utilisant la symétrie de la courbe par rapport à u = 0.

Table b. Fonction de répartition
Pour chaque valeur de u, la table donne la valeur de l'aire sous la courbe correspondante.

u	0,00	0,01	0,02	0,03	0,04	0,05
0,0	**0,5000**	0,5040	0,5080	0,5120	0,5160	0,5199
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987
0,3	0,6179	0,6217	**0,6255**	0,6293	0,6331	0,6368
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736

Exemple
• Pour u = 0 (0 + 0), la valeur de l'aire correspondante vaut 0,5 c'est-à-dire 50 % de la surface totale qui vaut 1. En pratique, cette aire représente la surface totale de -∞ à 0 inclus.
• Pour u = 0,32 (0,3 + 0,02), la valeur de l'aire correspondante vaut 0,6255 c'est-à-dire 62,55 % de la surface totale.

Voici quelques valeurs remarquables de la courbe à retenir :

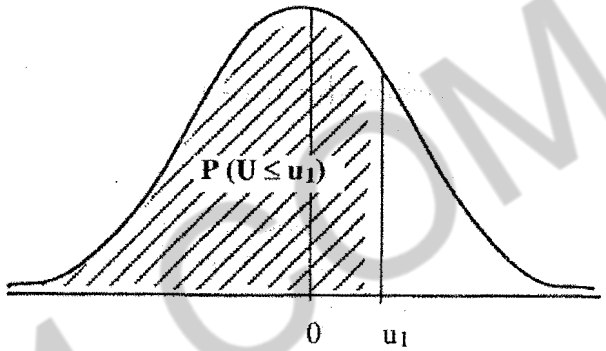
• u = 0, Φ(u) = 0,5
(la moitié de l'aire sous la courbe)

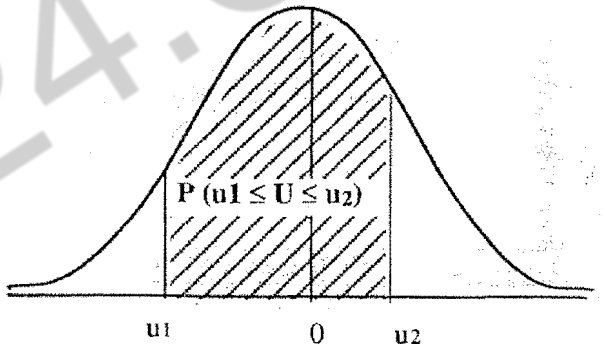
• u = 1,96, Φ(u) = 0,975

• u = 1, $\Phi(u) = 0{,}5 + \frac{0{,}683}{2} = 0{,}8413$

• u = 4,5, Φ(u) = 1
(totalité de la courbe)

En pratique, nous rencontrons les cas suivants :





Notez que :

$$\text{Prob}(U \le u_1) = \int_{\infty}^{0} f(u)du + \int_{0}^{u} f(u)du = 0{,}5 + \int_{0}^{u} f(u)du$$

3. Intervalle de confiance d'une moyenne

L'intervalle correspondant à (m – 2σ), (m + 2σ), par exemple, est appelé **« intervalle de confiance de la moyenne »** ou intervalle de pari au coefficient de sécurité de 95,5 %.
A cet intervalle de confiance s'attache un risque de (100 – 95,5) ≈ 5 %.

WWW.TALIB24.COM

La notion d'intervalle de confiance ainsi que sa signification concrète sera largement développée dans le chapitre suivant.

❑ Les notations et symboles des tables statistiques diffèrent, selon les auteurs, d'un ouvrage de statistique à l'autre.
Ainsi,

– la variable aléatoire est désignée par les lettres T, U ou Z et les valeurs numériques correspondantes par t, u ou z ;
– la fonction de répartition est désignée par F( ) ou f ( ), Φ ( ) ou π ( )

Vous ne devez pas pour autant vous perdre ! En effet, chaque table est en général accompagnée de commentaires qui expliquent le contenu.

❑ Certaines tables de la fonction de répartition donnent l'aire non pas de $-\infty$ à u, c'est-à-dire :

$$\int_{-\infty}^{u} f(u)\,du$$

mais l'aire de 0 à u, c'est-à-dire :

$$\int_{0}^{u} f(u)\,du$$

La table b ci-dessus est présentée alors ainsi :

Table b'

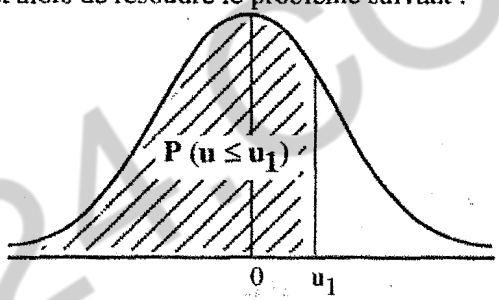
u	0,00	0,01	0,02	0,03	0,04	0,05
0,0	**0,000**	0,0040	0,0080	0,0120	0,0160	0,0199
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596

Compte tenu de la symétrie de la courbe, le passage de la table b' à la table b se fait tout simplement en ajoutant 0,5 à chaque probabilité.

4. A quoi sert la loi normale ?

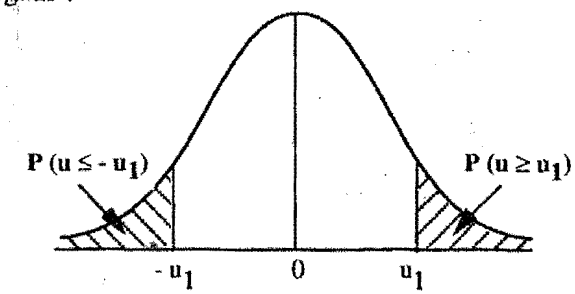
A. D'abord à rechercher des probabilités cumulées

Il faut **systématiquement** procéder au changement de variable, c'est-à-dire passer de x, la vraie variable, à u, variable centrée réduite par la relation $u = x - m/\sigma$.
La table I permet alors de résoudre le problème suivant :



Comment déterminer la probabilité que la variable **réduite** u appartienne à l'intervalle $[-\infty, u_1[$?
Cette table donne en effet la probabilité que u soit **inférieur ou égal** à une valeur donnée $\mathbf{u_1 > 0}$ (surface hachurée du graphe ci-dessus).

et si u est négatif ?



Si la valeur de u est négative, on voit sur le graphe ci-dessus que, la courbe étant symétrique par rapport à l'axe des ordonnées, les 2 surfaces hachurées sont égales.

Donc, $\text{Prob}(u \le -u_1) = \text{Prob}(u \ge u_1)$

WWW.TALIB24.COM

Or l'ensemble ($u > u_1$) est complémentaire de l'ensemble ($u \leq u_1$), puisque l'aire totale sous la courbe vaut 1.

Donc,

$\text{Prob}(u > u_1) = 1 - \text{Prob}\ (u \leq u_1)$

autrement dit,

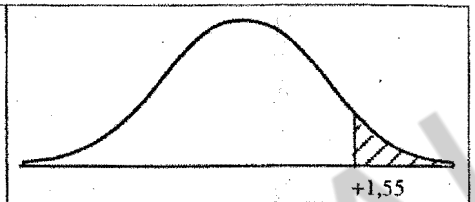
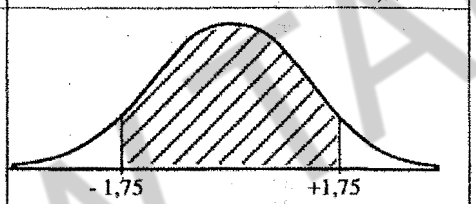
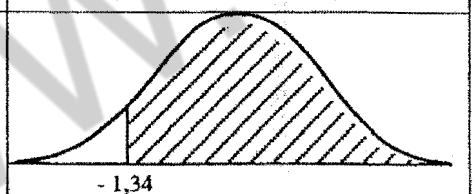
$$\textbf{Prob}(u \leq -u_1) = 1 - \textbf{Prob}(u \leq u_1)$$

Exercice

Quelle est la proportion des valeurs de u :

1. supérieures à 1,55 ?
2. inférieures en valeur absolue à 1,75 ?
3. supérieures à -1,34 ?

Corrigé

1. Prob(u > +1,55) = 1 – Prob(u ≤ 1,55) = 1 – 0,9394 = 0,06	
2. Prob(| u | < +1,75) Prob(–1,75 < u < +1,75) = Prob(u ≤ 1,75) – Prob(u ≤ –1,75) = 0,95994 – [1 - 0,95994] = 0,92	
3. Prob(u > –1,34) = 1 – Prob (u ≤ –1,34) = 1 – [1 – 0,90988] = 1 – 0,09012 = 0,91	

B. Intérêts de la distribution de la loi normale

La loi normale est souvent appelée « loi du hasard » pour la raison suivante : un événement dû au hasard peut être considéré comme le résultat d'un grand nombre de causes, indépendantes entre elles, exerçant chacune une petite influence, du même ordre de grandeur.

Soit, par exemple, X une variable aléatoire qui est égale à la somme de plusieurs autres :

$$X = X_1 + X_2 + \ldots + X_i + \ldots + X_n$$

Les variables X_i sont indépendantes entre elles ; elles obéissent à des lois de probabilités inconnues (quelconques) ; elles prennent cependant des valeurs du même ordre de grandeur. Si n est assez grand, la variable aléatoire X obéit à une loi normale.

Cet ensemble de conditions se trouve souvent réalisé dans les phénomènes naturels comme en biologie, agronomie etc. **C'est ce qui explique l'importance du rôle de la loi normale dans l'étude des distributions expérimentales.**

WWW.TALIB24.COM

III. Approximation d'une loi binomiale par une loi normale

Si la probabilité p de l'événement A est telle que **p est, relativement, différent de 0 ou 1**, et le nombre d'épreuves **n** est **assez grand** ($n \geq 30$) de sorte que le produit **n.p.q soit > 10**, la loi binomiale de paramètres n et p peut être approximée par une loi normale de paramètres np et $\sqrt{npq}$.
On passe alors de :

$$\text{Prob}(x) = \frac{n\,!}{x\,!\,(n-x)!}\; p^x\, q^{n-x} \qquad \text{avec } m = n.p$$

$$\text{à : } \text{Prob}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-m}{\sigma}\right]^2} = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}\left[\frac{x-n.p}{\sqrt{n.p.q}}\right]}$$

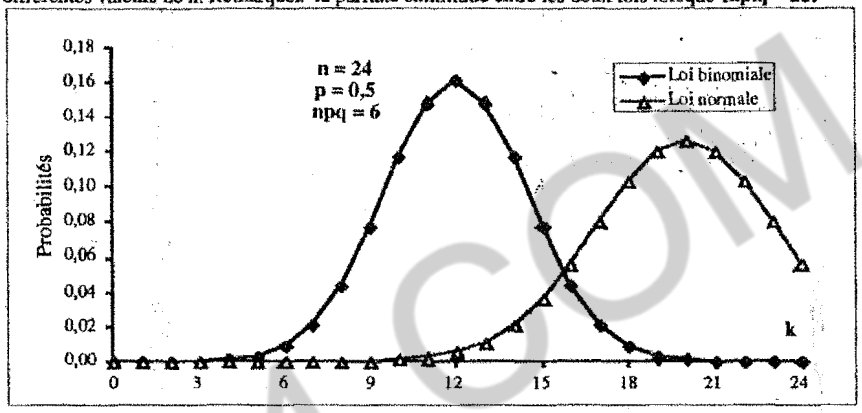
Cette approximation correspond au théorème de Moivre-Laplace.

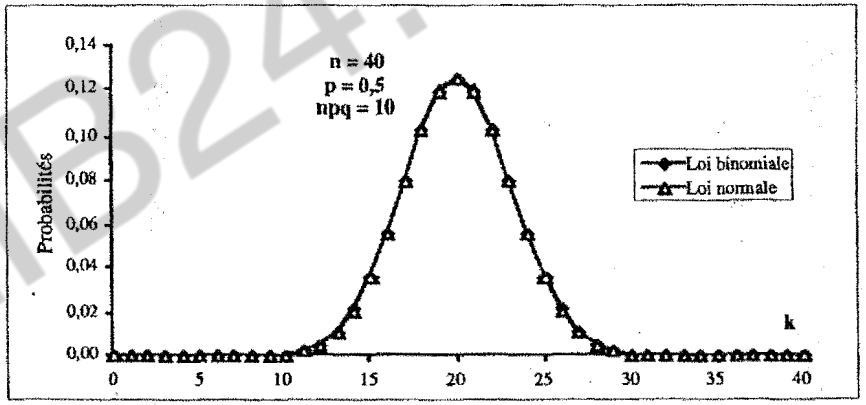
Si k_1 et k_2 sont deux entiers compris entre 0 et n, les intervalles $]k_1 , k_2[$ et $[k_1 , k_2]$ n'ont pas la même probabilité pour la loi binomiale, alors qu'ils ont la même probabilité dans le cas de la loi normale. Cette particularité est due au fait qu'on approxime – on dit aussi on approche – une loi relative à une variable **discrète** par une loi relative à une variable **continue.**
La correction consiste alors à remplacer :
- l'intervalle $]k_1, k_2[$ par $[k_1 + 0,5 , k_2 - 0,5]$;
- l'intervalle $[k_1, k_2]$ par $[k_1 - 0,5 , k_2 + 0,5]$.

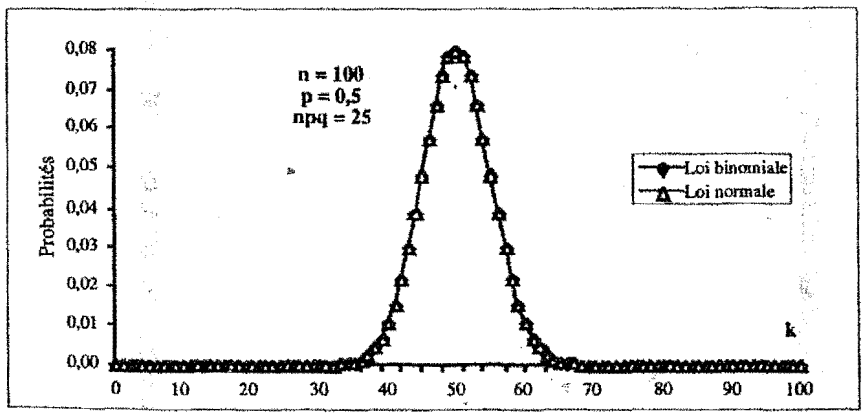
On appelle ce type de transformation « **correction de continuité** ».

Deux exemples d'illustrations graphiques de l'approximation d'un loi binomiale par une loi normale vous sont présentés dans les pages suivantes.

Une illustration graphique du concept de l'approximation vous est présentée ci-dessous pour **p = 0,5** et différentes valeurs de n. Remarquez la parfaite similitude entre les deux lois lorsque **n.p.q = 25.**

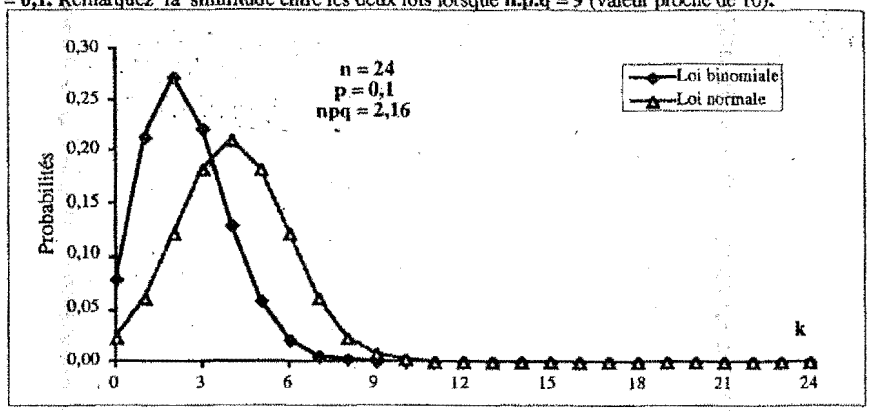


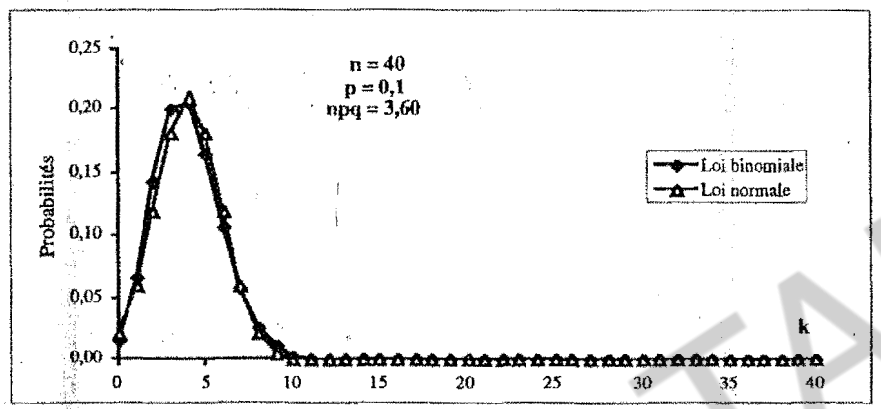


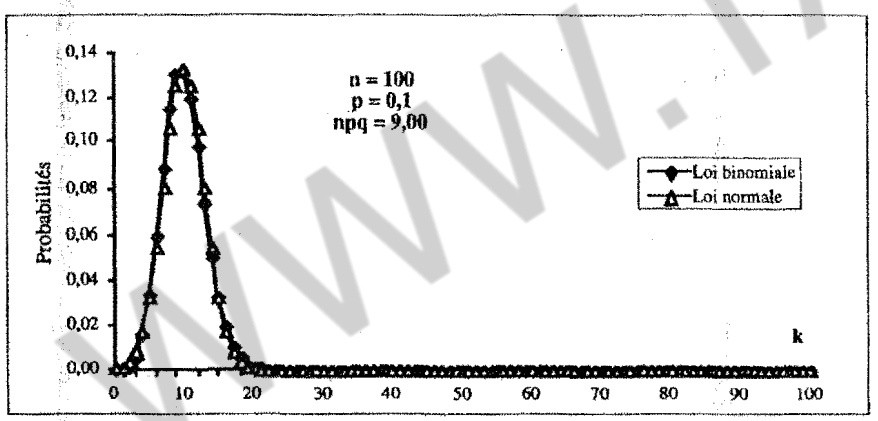


WWW.TALIB24.COM

Une illustration graphique du concept de l'approximation vous est présentée ci-dessous cette fois pour **p = 0,1.** Remarquez la similitude entre les deux lois lorsque **n.p.q = 9** (valeur proche de 10).







Exercice

On considère une population de bovins très nombreuse où chaque animal a une probabilité 0,2 d'être atteint par une mammite. On effectue 10 000 observations et on désigne par x la variable aléatoire égale au nombre d'animaux malades.

1. Quelle est la loi de probabilité de x ? Calculer le nombre moyen $\bar{x}$ d'animaux malades et l'écart type σ.

2. En déduire la probabilité d'avoir entre 1 900 et 2 040 animaux malades.

Corrigé

1. Il s'agit d'une loi binomiale dont les paramètres sont : n = 10 000 et p = 0,2

Donc $E(x) = \bar{x} = n \,.\, p = 10\,000 \,.\, 0{,}2 = 2\,000$

$\sigma(x) = \sqrt{n.\,p\,.\,q} = \sqrt{10\,000 \,.\, 0{,}2 \,.\, 0{,}8} = 40$

2. Cette loi binomiale $\mathcal{B}(n, p)$ peut être approximée par une loi normale $\mathcal{N}(2\,000, 40)$.

Chercher la probabilité Prob(1 900 < x < 2 040), revient à chercher, en tenant compte de la correction de continuité :

$$\text{Prob}(1\,900 < x < 2\,040) = \text{Prob}(1\,899{,}5 < x < 2\,040{,}5)$$

$$\text{Prob}\left[\frac{1\,899{,}5 - 2\,000}{40} < u < \frac{2\,040{,}5 - 2\,000}{40}\right]$$

$$= \text{Prob}(-2{,}51 < u < 1{,}01) = \text{Prob}(u < 1{,}01) - [\text{Prob}(u < -2{,}51)]$$

$$= 0{,}8438 - (1 - 0{,}9940) = \mathbf{0{,}8377}$$

WWW.TALIB24.COM

Testez vos connaissances

Exercice 1

Un échantillon de 80 veaux est extrait d'une population dont la variable aléatoire « Poids » suit une loi normale d'espérance mathématique 500 kg et d'écart type 40 kg.

1. Dans cet échantillon, combien doit-on s'attendre à trouver de veaux pesant :
a. Plus de 560 kg ?
b. Moins de 480 kg ?
c. Entre 450 kg et 550 kg ?

2. On sélectionne, pour la reproduction, les 15% supérieurs de l'échantillon. A partir de quel poids un animal sera-t-il sélectionné ?

Exercice 2

Pour une population de candidats à la filière de BTS.PA la distribution des notes dans un test est normale avec une moyenne $\bar{x} = 32{,}3$ et un écart type $\sigma = 8{,}5$.
On décide que 10 % des sujets seront orientés en « Ingéniorat » parce que leur niveau est très haut, et que 30 % des sujets seront orientés en BTA parce que leur niveau est trop bas.

WWW.TALIB24.COM

Entre quelles limites la note d'un sujet devra-t-elle se placer pour qu'il soit admis dans la filière de BTS.PA.

Exercice 3

On considère deux portées de porcelets d'une race donnée ; la première portée est de 7 porcelets et la deuxième de 5. On regroupe tous les porcelets. Tous les porcelets ont même apparence et ont la même probabilité d'être choisis.

1. On prélève, l'un après l'autre, avec remise et de façon indépendante, 3 porcelets. On appelle X la variable aléatoire prenant pour valeur le nombre de porcelets de la première portée figurant parmi les 3 porcelets tirés.

a. Quelle est la loi de probabilité de la variable aléatoire X ?

b. Calculer la probabilité pour qu'il y ait au moins un porcelet de la première portée.

c. Donner l'espérance et la variance de X.

2. On prélève simultanément 3 porcelets. On appelle Y la variable aléatoire prenant pour valeur le nombre de porcelets de la première portée figurant parmi les 3 porcelets tirés. Déterminer la loi de probabilité de Y ; calculer son espérance et sa variance.

Exercice 4

On considère une population très nombreuse où chaque individu est susceptible de posséder un caractère A avec une probabilité de 0,2.

On effectue 10 000 observations et on désigne par X la v.a. égale au nombre d'individus ayant le caractère A.

1. Indiquer la loi de probabilité de X et calculer son espérance mathématique E(X) ainsi que son écart type σ_X.

2. On approche cette loi de probabilité par une loi classique. Dans ces conditions, calculer les probabilités des événements suivants :

$1900 < X < 2040$; $X < 1950$; $X > 2030$

Vérifiez vos réponses à la page suivante

WWW.TALIB24.COM

Réponses au test

Exercice 1

1.

a. $\text{Prob}(X > 560) = \text{Prob}(U > \frac{560 - 500}{40})$

$\text{Prob}(U > 1{,}5) = 1 - \Phi(1{,}5) = 1 - 0{,}9332 = 0{,}0668$

soit sur 80 veaux $\Rightarrow$ 5 veaux

b. $\text{Prob}(X < 480) = \text{Prob}(U < \frac{480 - 500}{40}) = \text{Prob}(U < -0{,}5) = \Phi(-1{,}5)$

$= 1 - \Phi(0{,}5) = 1 - 0{,}6915 = 0{,}3085$

soit sur 80 veaux $\Rightarrow$ 25 veaux

c. $\text{Prob}(450 < X < 550) = \text{Prob}(\frac{450 - 500}{40}) < U < \frac{550 - 500}{40}) =$

$= \text{Prob}(-1{,}25 < U < 1{,}25)$

$= \Phi(1{,}25) - \Phi(-1{,}25) = 2\,\Phi(1{,}25) - 1 = 2 \times 0{,}8944 - 1 = 0{,}7888$

soit sur 80 veaux $\Rightarrow$ 63 veaux

2. $\text{Prob}(X > x_m) = 0{,}15$ ou $\text{Prob}(X < x_m) = 0{,}85$

$\text{Prob}(U < \frac{x_m - 500}{40}) = 0{,}85$

$\Phi(\frac{x_m - 500}{40}) = 0{,}85$ d'où $\frac{x_m - 500}{40} = 1{,}04$

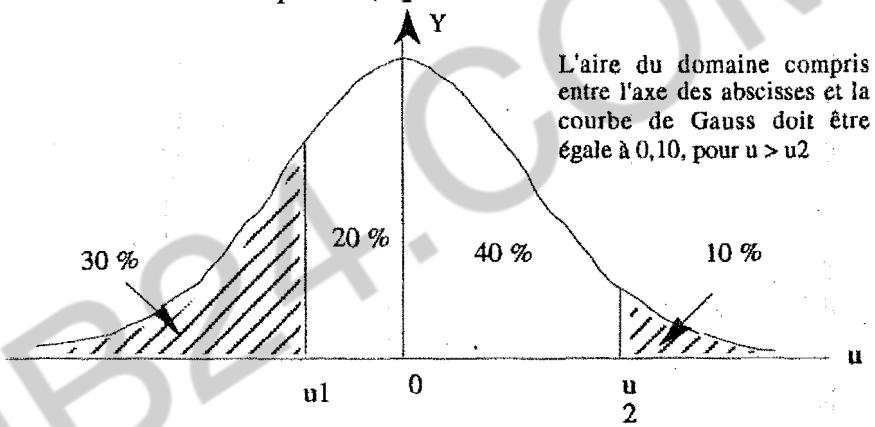
et

$x_m = 541{,}6$ kg

Exercice 2

$u = \frac{x - \bar{x}}{\sigma}$

1. Calcul de la limite supérieure, u_2 de u



$\Phi(x \le u_2) = 1 - 0{,}1 = 0{,}9 \quad \Rightarrow \quad u_2 = 1{,}28$ (voir table I)

2. Calcul de la limite inférieure de u : u_1

$\Phi(x \ge u_1) = 1 - \Phi(x \le u_1) = 0{,}20 + 0{,}5 = 0{,}70$

la table I donne la valeur de $-u_1 = 0{,}52$

$u_1 = -0{,}52$

$u_2 = 1{,}28 \quad \Rightarrow \quad -0{,}52 \le \frac{x - 32{,}3}{8{,}5} \le 1{,}28$

$x = (-0{,}52 \,.\, 8{,}5) + 32{,}3$ (limite inférieure)

et $x = (1{,}28 \,.\, 8{,}5) + 32{,}3$ (limite supérieure)

$$\mathbf{27{,}9 \le x \le 43{,}2}$$

WWW.TALIB24.COM

Exercice 3

1.

a. L'expérience est répétée n fois ($n = 3$) dans des conditions identiques et indépendantes et qui donnent lieu à deux résultats qui s'excluent mutuellement, ou le porcelet appartient à la 1re portée ou il appartient à la 2e. La v. a X est donc une v. a binomiale de paramètres :

$n = 3,\ p = \frac{7}{12}$

Sa loi de probabilité est : $P(x = k) = C_3^k (\frac{7}{12})^k (\frac{5}{12})^{3-k}$

$0 \le k \le 3$

b. $1 - P(x = 0) = 1 - (\frac{5}{12})^3 = 0{,}9271$

c. $E(X) = np = 1{,}75$

$V(X) = npq = 0{,}7292$

2. La loi de probabilité de Y n'est pas connue *a priori*, il faut faire un tableau ou à chaque valeur Yi de Y on associe la probabilité P_{Yi} correspondante :

Yi	0	1	2	3	
P_{Yi}	$\frac{C_7^0 \times C_5^3}{C_{12}^3} = \frac{10}{220}$	$\frac{C_7^1 \times C_5^2}{220} = \frac{70}{220}$	$\frac{C_7^2 \times C_5^1}{220} = \frac{105}{220}$	$\frac{C_7^3 \times C_5^0}{220} = \frac{35}{220}$	
$Yi P_{Yi}$	0	$\frac{70}{220}$	$\frac{210}{220}$	$\frac{105}{220}$	$\Sigma Yi P_{Yi} = \mu = 1{,}75$
$Yi^2 P_{Yi}$	0	$\frac{70}{220}$	$\frac{420}{220}$	$\frac{315}{220}$	$\Sigma = \frac{805}{220}$
$(Yi - \mu)^2 P_{Yi}$	0,1392	0,1790	0,0298	0,2486	$\Sigma = 0{,}5966$

$V(Y) = \sum_i Yi^2 P_{Yi} - \mu^2 = 3{,}6591 - (1{,}75)^2 = 0{,}5966$

ou $V(Y) = \sum_i (Yi - \mu)^2 P_{Yi} = 0{,}5966$

Exercice 4

1. X suit une loi binomiale. En effet, chaque individu est indépendant des autres pour la possession du caractère A et l'on considère deux possibilités : avoir ou ne pas avoir le caractère A.

Les paramètres de la loi binomiale sont :

n = 10 000, nombre d'observations

p = 0,2, probabilité d'avoir le caractère A $\Rightarrow$ q = 1 – p = **0,8**

Dans ces conditions :

$E(X) = n.p = 10\,000 \times 0{,}2 = 2\,000$

$\sigma = \sqrt{n.p.q} = \sqrt{10\,000 \times 0{,}2 \times 0{,}8} = 40$

2. Cette loi binomiale peut être approchée par une loi normale de moyenne 2 000 et d'écart type 40.

Soit : $u = \frac{X - m}{\sigma} \Rightarrow X = u.\sigma + m$

• **1 900 < X < 2 040**

$= \text{Prob}\left(\frac{1\,900 - m}{\sigma} < u < \frac{2\,040 - m}{\sigma}\right)$

$= \text{Prob}\left(\frac{1\,900 - 2\,000}{40} < u < \frac{2\,040 - 2\,000}{40}\right)$

$= \text{Prob}(-2{,}5 < u < 1) = \text{Prob}(u < 1) - \text{Prob}(u < -2{,}5)$

$= \text{Prob}(u < 1) - [1 - \text{Prob}(u < 2{,}5)] = 0{,}83134 - (1 - 0{,}99379)$

$= \mathbf{0{,}835\ 13}$

WWW.TALIB24.COM

• **X < 1 950**

Prob(X < 1 950)

$$= \text{Prob}(m + u\sigma < 1\,950)$$

$$= \text{Prob}\left(u < \frac{1\,950 - 2\,000}{40}\right)$$

$$= \text{Prob}(u < -1{,}25) = 1 - [\text{Prob}\ (u < 1{,}25)] = 1 - 0{,}89435$$

$$= \mathbf{0{,}105\ 65}$$

• **X > 2 030**

$$= \text{Prob}\left(u > \frac{2\,030 - 2\,000}{40}\right)$$

$$= \text{Prob}(u > 0{,}75) = 1 - [\text{Prob}\ (u < 0{,}75)] = 1 - 0{,}77337$$

$$= \mathbf{0{,}226\ 63}$$

Exercices supplémentaires

Exercice 1

On a mesuré le poids de 3 000 moutons et on a trouvé une moyenne de 65 kg et un écart type de 3 kg.

En supposant que la variable « poids » suit une distribution normale, déterminer le nombre de moutons ayant un poids compris entre 60 et 70 kg.

Exercice 2

On considère un échantillon de 10 veaux soumis à l'engraissement pour lesquels on a étudié le gain moyen quotidien (GMQ). On a obtenu une moyenne de 902,9 g/j et un écart type de 30,54 g/j.

1. Donner une estimation ponctuelle de l'écart type du GMQ de la population d'où a été extrait cet échantillon.

On donne l''écart type estimé de la population $\hat{\sigma} = \sqrt{\frac{n}{n-1}}\,s$

2. Soit X la variable aléatoire égale au GMQ ; on suppose qu'elle suit une loi normale. Calculer p(880 < X < 950).

3. On sélectionne les 20% supérieurs de la population. A partir de quel GMQ un veau sera-t-il sélectionné ?

WWW.TALIB24.COM

Exercice 3

On a constaté que la répartition du taux de cholestérol pour un grand nombre de personnes est la suivante :
- Taux inférieur à 165 cg : 58 %
- Taux compris entre 165 et 180 cg : 38 %
- Taux supérieur à 180 cg : 4 %

1. Sachant que la répartition suit une loi normale, calculer la valeur moyenne du taux de cholestérol et l'écart type moyen.

2. On admet que les personnes dont le taux est supérieur à 183 cg doivent subir un traitement. Quel est le nombre de personnes à soigner dans une population de 100 000 personnes ?

Exercice n°4

D'une grande population d'animaux de la même race, dont 64% ne sont pas atteints d'une maladie A, on tire au hasard 16 d'entre eux.*(on assimile ce tirage à un tirage avec remise)*. Soit X la variable aléatoire prenant pour valeur le nombre d'animaux n'ayant pas cette maladie :

1. Quelle est la loi de X ? Calculer son espérance E(X) et sa variance V(X).

2.

a. Quelle est la probabilité d'avoir 10 d'entre eux non atteints ?

b. Quelle est la probabilité d'avoir au plus 15 d'entre eux non atteints ?

3. En utilisant une approximation que l'on justifiera, déterminer

$$p(6 \leq X \leq 11)$$

Corrigé des exercices supplémentaires

Exercice 1

Soit x = variable « poids »

m = 65 kg

σ = 3 kg

La proportion de moutons ayant un poids compris entre x_1 = 60 kg et x_2 = 70 kg est donnée par la loi normale.

avec $u = \frac{x-m}{\sigma}$

$$u_1 = \frac{x_1-m}{\sigma} = \frac{60-65}{3} = -1{,}67$$

$$u_2 = \frac{x_2-m}{\sigma} = \frac{70-65}{3} = +1{,}67$$

$$\text{Prob}(x_1 \leq x \leq x_2) = \int_{u_1}^{u_2} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$= \Phi(1{,}67) - \Phi(-1{,}67)$

$= 0{,}95254 - \{1 - \Phi(1{,}67)$ (table I)

$= 0{,}95254 - (0{,}04746) = 0{,}9051$

D'où le nombre de moutons ayant un poids compris entre 60 et 70 kg :

$3000 \times 0{,}9051 =$ **2715 moutons**

WWW.TALIB24.COM

Exercice 2

1.
$\bar{x} = 902,9 g/j$ est une bonne estimation de μ moyenne (*espérance*) de la population.

L'écart type estimé de la population $\hat{\sigma} = \sqrt{\frac{n}{n-1}}\, s$

d'où $\hat{\sigma} = \sqrt{\frac{10}{9}} \times 30,54 = 32,192\, g/j$

2.
$$= P\left(\frac{880-902,9}{32,192} < \frac{X-902,9}{32,192} < \frac{950-902,9}{32,192}\right) = 0,6890$$

3. On cherche la valeur x_m à partir de laquelle un veau peut être sélectionné, la réduction de cette valeur correspond à u tel que $\Phi(u) = 0,8$
d'où $u = 0,8416$

$$u = \frac{x_m - 902,9}{32,192} \quad \text{d'où :} \quad x_m = 929,99 \approx 930 \text{ g/j}$$

Exercice 3

1. Si la répartition suit une loi normale (Table I, fonction de répartition), on lit u dans la table I à partir de la fréquence cumulée :
$u = 0,2019$ correspond à la fréquence cumulée 0,58 et à 165 cg
$u = 1,75$ correspond à la fréquence cumulée 0,96 et à 180 cg
On a les 2 relations :

$$\frac{165-\bar{x}}{\sigma} = 0,20 \quad ; \quad \frac{180-\bar{x}}{\sigma} = 1,75$$

d'où :

$$165 - \bar{x} = 0,20\sigma$$
$$180 - \bar{x} = 1,75\sigma$$

c'est-à-dire, en retranchant membre à membre ces deux équations :

$1,55\,\sigma = 15 \quad \Rightarrow \quad \sigma = 15/1,55 = \mathbf{9,68\ cg}$

et $\bar{x} = 165 - 0,20\,\sigma = \mathbf{163\ cg}$

2.
A 183 cg correspond $u = \frac{183-162}{9,68} = 2,07$

donc, 98,08 % (table I) de la population a une valeur inférieure à 183 ; autrement dit, 1,92 % doivent subir un traitement, soit 1 920 dans une population de 100 000.

Exercice 4

1. Le tirage est effectué 16 fois dans des conditions identiques et indépendantes, chaque tirage donne lieu a deux résultats possibles qui s'excluent mutuellement, (ou ils sont atteints de la maladie A ou ils ne sont pas atteints). La variable aléatoire X est donc une v.a binomiale de paramètres n = 16 et p = 0,64
Sa loi de probabilité est donc :
$P(X = k) = C_{16}^{k}(0,64)^{k}(0,36)^{16-k)}$
$0 \le k \le 16$
$E(X) = np = 16 \times 0,64 = 10,24$
$V(X) = npq = 16 \times 0,64\ 0,36 = 3,6864$

2.
a. $P(X = 10) = C_{16}^{10}(0,64)^{10} \times (0,36)^{6} = 0,2010$

b. P(*au plus 15 non atteints*) = 1 – P(*16 non atteints*)
$= 1 - C_{16}^{16}(0,64)^{16}(0,36)^{0}$
$= 1 - (0,64)^{16} = 1 - 0,0008 = 0,9992$

WWW.TALIB24.COM

3. On peut approximer la v.a X binomiale, à une variable normale de même espérance et de même variance. Il faut pour cela faire la correction de continuité qui transforme la variable binomiale discontinue en une v.a normale continue.

$$P(6 \le X \le 11) = P(5,5 < X < 11,5)$$

$$P(6,5 < X < 11,5) = P\left(\frac{6,5-10,24}{\sqrt{3,6864}} < \frac{X-10,24}{\sqrt{3,6864}} < \frac{11,5-10,24}{\sqrt{3,6864}}\right) = 0,7198$$

Table I

Table de la distribution normale réduite

FONCTION DE RÉPARTITION

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-1/2\, u^2}\, du$$

Exemple : $\Phi(0,52) = 0,6985$; $\Phi(-1,93) = 1 - \Phi(1,93) = 1 - 0,97320 = 0,02680$

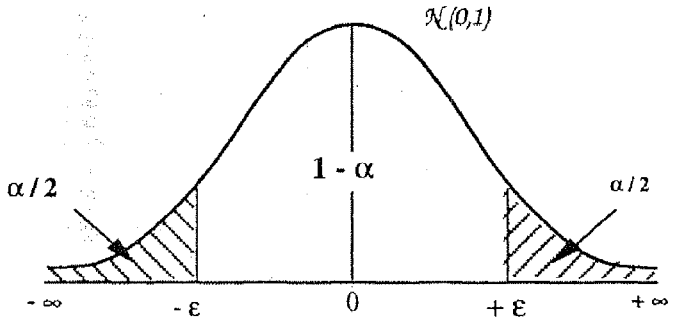
u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91308	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

WWW.TALIB24.COM

Table II

Table de la loi normale centrée, réduite $\mathcal{N}(0,1)$

OU TABLE DE L'ÉCART RÉDUIT



La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour ε = 1,96, la probabilité est α = 0,00 + 0,05 = 0,05

α	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	**2,577**	2,327	2,171	2,054	**1,960**	1,881	1,812	1,751	1,696
0,10	1,645	**1,598**	1,555	1,514	1,476	**1,440**	1,405	1,372	1,341	1,311
0,20	1,282	**1,254**	1,227	1,201	1,175	**1,150**	1,127	1,103	1,080	1,058
0,30	1,037	**1,015**	0,995	0,974	0,954	**0,935**	0,915	0,897	0,878	0,860
0,40	0,842	**0,824**	0,806	0,789	0,772	**0,755**	0,739	0,723	0,706	0,690
0,50	0,675	**0,659**	0,643	0,628	0,613	**0,598**	0,583	0,568	0,553	0,539
0,60	0,524	**0,510**	0,496	0,482	0,468	**0,454**	0,440	0,426	0,412	0,399
0,70	0,385	**0,372**	0,358	0,345	0,332	**0,319**	0,305	0,292	0,279	0,266
0,80	0,253	**0,240**	0,228	0,215	0,202	**0,189**	0,176	0,164	0,151	0,138
0,90	0,126	**0,113**	0,100	0,088	0,075	**0,063**	0,050	0,038	0,025	0,013

Tables pour les petites valeurs de α

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
ε	3, 290 53	3,890 59	4,417 17	4,891 64	5,326 72	5,730 73	6,109 41

Annexe : la loi normale, reine des statistiques

extrait de la Revue *Pour la Science*
numéro hors série - avril 1996

WWW.TALIB24.COM

19 Échantillons

I Objectif

L'estimation statistique consiste à obtenir des informations sur des variables liées à une population à partir d'un échantillon. En statistique descriptive un échantillon est défini comme un sous-ensemble de la population, en théorie des probabilités un échantillon est constitué de variables aléatoires qui suivent une même loi. Les résultats issus de l'échantillon permettent d'induire un certain nombre de caractéristiques sur la loi d'une variable aléatoire.

II L'essentiel à savoir

• Définition

$(X_1, X_2, \ldots, X_i, \ldots, X_n)$ est un n-échantillon d'une variable aléatoire X si les variables aléatoires X_i sont indépendantes et suivent la même loi que X.

$(x_1, x_2, \ldots, x_i, \ldots, x_n)$ est une réalisation du n-échantillon de X après n expériences aléatoires.

Une autre série d'expériences donnera une autre réalisation $(x'_1, x'_2, \ldots, x'_i, \ldots, x'_n)$.

→ ***Remarque***

Le vocabulaire employé est souvent le même pour désigner un n-échantillon ou une réalisation de ce n-échantillon. Ces définitions permettent de formaliser la théorie de l'estimation.

EXEMPLE

Les pièces d'une fabrication ont une probabilité p d'être défectueuses. À la $i^{\text{ème}}$ expérience aléatoire qui consiste à prélever au hasard une pièce dans la fabrication on associe la variable X_i de Bernoulli qui prend la valeur 1 si la pièce n° i est défectueuse et la valeur 0 si la pièce n° i n'est pas défectueuse. Supposons que n soit égal à 20, une réalisation de ce 20-échantillon pourrait être :

$$(1,0,0,0,0,0,1,0,0,0,0,1,0,0,0,1,0,0,0,0)$$

$x_1 = 1$. Ce qui veut dire que la première pièce prélevée est défectueuse, les cinq suivantes ne le sont pas : $x_2 = x_3 = x_4 = x_5 = x_6 = 0$, puis $x_7 = 1$, etc.

La proportion de pièces défectueuses dans l'échantillon est :

$$p_0 = \overline{x} = \frac{x_1 + x_2 + \ldots + x_{20}}{20} = \frac{\sum x_i}{20} = \frac{4}{20} = 0,2$$

Une réalisation de l'échantillon permet d'estimer p par $p_0 = 0,2$.

III Compléments : simulation d'un échantillon

Pour de nombreuses applications pratiques il peut être utile de simuler un échantillon d'une loi de probabilité connue.

A. Simulation d'un échantillon d'une loi de Bernoulli

Commençons par le cas le plus simple consistant à simuler une loi de Bernoulli de paramètre $p = 0,1$. Il est possible de simuler des expériences qui résultent de cette loi en mettant dans une urne une boule blanche et neuf boules noires, à chaque tirage avec remise d'une boule blanche est associé $x = 1$. Une autre façon consiste à utiliser des tables de nombres au hasard. Ces tables fournissent des suites de chiffres compris entre 0 et 9, chaque chiffre a une probabilité 1/10 d'apparaître, c'est la réalisation d'une loi discrète uniforme et les réalisations sont indépendantes.

EXEMPLE

Soit l'extrait suivant d'une table de nombres au hasard :

339 103 487 925 745 807 636 10

Cet extrait conduit à la réalisation suivante d'un échantillon de Bernoulli de paramètre $p = 0,1$:

$$0,0,0,1,0,0,0,0,0\ldots1,0$$

Chaque fois que le chiffre 1 apparaît X prend la valeur 1 et X prend la valeur 0 sinon. En prenant les chiffres deux à deux consécutifs, il apparaît aléatoirement des nombres entiers entre 00 et 99 de façon uniforme avec une probabilité de 1/100.

Ainsi, par exemple $P(X = 27) = P(2).P(7) = \frac{1}{10} \times \frac{1}{10} = \frac{1}{100}$.

Pour simuler une loi de Bernoulli de paramètre $p = 0,24$, il suffit de prendre les chiffres deux par deux, si le nombre est compris entre 00 et 23 inclus, on décide que $x = 1$ et si le nombre est compris entre 24 et 99, on décide $x = 0$.

Pour simuler une loi Binomiale de paramètre n et p, il suffit d'ajouter des suites de n variables de Bernoulli indépendantes de paramètre p.

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

B. Simulation d'une loi quelconque continue

• Théorème

Soit X une variable aléatoire continue et $F(x) = P(X < x)$ alors $Y = F(X)$ est une variable aléatoire uniforme et continue sur l'intervalle $[0, 1]$.

• Démonstration du théorème

Soit G la fonction de répartition de Y.

$$G(y) = P(Y < y) = P(F(X) < F(x)) = P(X < x) = F(x) = y$$

Puisque $G(y) = y$, la distribution de Y est uniforme sur $[0, 1]$.
C'est la monotonie de la fonction F qui a permis d'écrire :

$$X < x \Leftrightarrow F(X) < F(x) \text{ et } P(X < x) = P[F(X) < F(x)]$$

• Obtention d'une loi quelconque à partir de nombres au hasard

Soit X une variable aléatoire et F sa fonction de répartition.
Pour obtenir $(x_1, x_2, \ldots, x_i, \ldots, x_n)$, réalisation d'un n-échantillon de X, il faut déterminer d'abord n nombres au hasard indépendants $y_1, y_2, \ldots, y_i, \ldots, y_n$, ces nombres étant compris entre 0 et 1. Et puisque F est bijective :

$$\begin{aligned} y_1 = F(x_1) &\iff x_1 = F^{-1}(y_1) \\ y_2 = F(x_2) &\iff x_2 = F^{-1}(y2_1) \\ &\vdots \\ y_n = F(x_n) &\iff x_n = F^{-1}(y_n) \end{aligned}$$

• Obtention de nombres au hasard entre 0 et 1

Les nombres $y_1, y_2, \ldots, y_i, \ldots, y_n$ s'obtiennent aisément à partir d'une table de nombres au hasard ou encore ils sont générés automatiquement par une calculatrice ou un ordinateur.
Supposons que nous voulions des nombres compris entre 0 et 1 avec trois chiffres significatifs après la virgule, il suffit de prendre les chiffres de la table trois par trois et de diviser par mille.

EXEMPLE

339 103 487 925 745 extrait de la table conduit à $y_1 = 0,339$; $y_2 = 0,103$; $y_3 = 0,487$; ...
Si l'on souhaite quatre chiffres significatifs après la virgule, il suffit de prendre les chiffres 4 par 4 et de les diviser par 10 000. Ce qui donne $y_1 = 0,3391$; $y_2 = 0,0348$; $y_3 = 0,7925$; ...

IV Application

ÉNONCÉ

1. Simuler un échantillon de taille 5 d'une loi normale centrée réduite.
2. Simuler un échantillon de taille 5 d'une loi normale de moyenne 175 et d'écart type 5.

SOLUTION

1. Si $L(T) = N(0, 1)$, à partir d'une table de nombres au hasard, on simule un échantillon de taille 5 d'une loi uniforme sur [0, 1]. Trois chiffres significatifs après la virgule constituent une approximation suffisante, aussi les chiffres de la table sont extraits 3 par 3

$$y_1 = 0,339 \ ; \ y_2 = 0,103 \ ; \ y_3 = 0,487 \ ; \ y_4 = 0,925 \ ; \ y_5 = 0,745.$$

F étant la fonction de répartition de la loi $N(0, 1)$,

$$y_1 = F(t_1) \Leftrightarrow t_1 = F^{-1}(y_1)$$

$$y_1 = F(t_1) = P(T < t_1) = \int_{-\infty}^{t_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = 0,339$$

F^{-1} n'ayant pas d'expression algébrique simple, t_1 se détermine à l'aide de la table statistique de la loi $N(0, 1)$, soit $t_1 = -0,415$ et ainsi de suite :

$$t_2 = -1,267 \ ; \ t_3 = -0,032 \ ; \ t_4 = 1,440 \ ; \ t_5 = 0,657.$$

2. $L(X) = N(170, 5)$, $\quad T = \dfrac{X - 170}{5}$, $\quad L(T) = N(0, 1)$

D'où $X = 5T + 170$ et pour chaque expérience $x = 5t + 170$

$t_1 = -0,415$ d'où $x_1 = 5(-0,415) + 170 = 167,92$
$t_2 = -1,267$ d'où $x_2 = 5(-1,267) + 170 = 163,66$ etc.

Table

Nombres au hasard

	5	10	15	20	25	30	35	40	45	50
	33910	34879	25745	80763	61033	56033	78902	47008	72488	57949
	50230	63237	94083	93634	71652	02656	57532	60307	91619	48916
5	84980	62458	09703	78397	66179	46982	67619	39254	90763	74056
	22116	33646	17545	31321	65772	86506	09811	82848	92211	51178
	68645	15068	56898	87201	40115	27524	42221	88293	67592	06430

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

20 Estimateur et estimation ponctuelle

I Objectifs

• L'étude de la distribution et des caractéristiques d'une ou plusieurs variables concernant une population ne peut généralement être faite sur l'ensemble des individus qui composent la population. Par exemple, pour connaître les intentions de vote d'une population qui comporte des millions d'électeurs, il est impossible de procéder, hormis le jour des élections, autrement que par un sondage c'est-à-dire par le recueil des informations auprès d'un échantillon issu de cette population.

• L'objectif de cette fiche est de présenter les principales fonctions aléatoires qui permettent d'estimer une moyenne, une variance et une proportion, et surtout de donner les propriétés de ces fonctions qui sont appelées estimateurs. Ce sont les propriétés de ces estimateurs qui permettent de déterminer une marge de l'erreur sous forme d'intervalles de confiance lorsqu'une caractéristique d'une population est estimée sur un échantillon.

II L'essentiel à savoir

A. Estimateur et estimation

Un n-échantillon d'une variable aléatoire X est une suite de n variables indépendantes et de même loi que X.

• Définition d'un estimateur

Un estimateur d'un paramètre θ dont dépend la loi de probabilité d'une variable aléatoire X est une fonction d'un n-échantillon de X $(X_1, X_2, \ldots, X_i, X_n)$

$$\boxed{\Theta_n = T(X_1, X_2, \ldots, X_i, \ldots, X_n)}$$

© Dunod – La photocopie non autorisée est un délit.

L'estimateur est une fonction de variables aléatoires X_i, c'est aussi une variable aléatoire, pour cette raison l'estimateur est noté avec des lettres majuscules :

$$\Theta_n \text{ ou } T_n \text{ ou } T \text{ ou encore } \widehat{\Theta}$$

• Choix de l'estimateur-estimation

Il existe une infinité de façons de choisir un estimateur de θ. Toutefois il est judicieux de choisir un estimateur qui possède de bonnes propriétés statistiques (voir ces propriétés au paragraphe suivant). Dans une première approche on peut retenir l'idée qu'après « expériences aléatoires » il est associé à $(x_1, x_2, \ldots, x_i, x_n)$ réalisation d'un n-échantillon de X, une valeur $T(x_1, x_2, \ldots, x_i, x_n) = \widehat{\theta}_n$. Cette valeur numérique $\widehat{\theta}_n$ est l'estimation la plus appropriée pour le paramètre θ. C'est-à-dire que $\widehat{\theta}_n$ est en quelque sorte la meilleure valeur approchée de θ.

EXEMPLE

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_i + \ldots + x_n}{n} = \frac{\sum x_i}{n} = T(x_1, x_2, \ldots, x_i, x_n)$$

La moyenne de l'échantillon $\overline{x}$ est un choix intuitif, pour choisir une estimation de $m = E(X)$, l'estimateur en découle immédiatement, c'est

$$\overline{X} = T(X_1, X_2, \ldots, X_i, X_n) = \frac{\sum X_i}{n}.$$

Ce qui suit confirme ce choix, en ce sens que cet estimateur a de bonnes propriétés statistiques.

L'estimation est une réalisation de la variable aléatoire T_n estimateur de θ ; pour cette raison une estimation est une valeur numérique souvent notée en minuscule t_n ou $\widehat{\theta}$.

B. Propriétés souhaitables des estimateurs

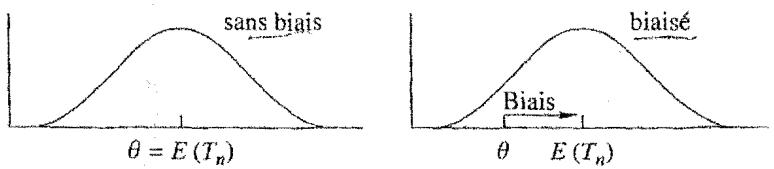
• Définition d'un estimateur sans biais ou sans distorsion

T_n est un estimateur sans biais de θ si :

$$\boxed{E(T_n) = \theta}$$

θ est la valeur exacte mais inconnue du paramètre à estimer.

Si $E(T_n) \neq \theta$, l'estimateur est biaisé, le biais est $E(T_n) - \theta$.



WWW.TALIB24.COM

EXEMPLE

$T_n = \overline{X} = \dfrac{\sum X_i}{n}$ est un estimateur sans biais de $m = E(\overline{X})$, en effet :

$$E(\overline{X}) = E\left(\frac{\sum X_i}{n}\right) = \frac{1}{n}E\left(\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{nE(X)}{n} = E(X) = m$$

Remarque

Si $E(T_n)$ tend vers θ quand n tend vers l'infini, l'estimateur T_n de θ est asymptotiquement sans biais.

• Convergence en probabilité

T_n converge en probabilité vers θ si :

$$\left[P(\theta - \varepsilon < T_n < \theta + \varepsilon)\right] \to 0 \text{ quand } n \to \infty$$

À un niveau de connaissances pas trop approfondies, il suffit de savoir que si les deux conditions suivantes sont remplies l'estimateur T_n converge en probabilité vers θ.

1) $E(T_n) = \theta$ (ou $E(T_n) \to \theta$) 2) $V(T_n) \to 0$ quand $n \to \infty$

C'est le théorème de convergence en probabilité, proposé dans la fiche n °17.

EXEMPLE

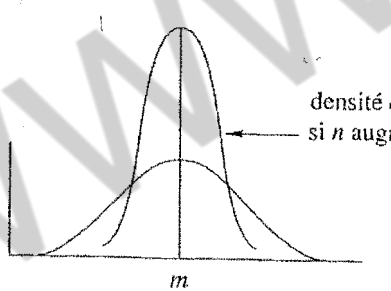
$\overline{X}$ est un estimateur convergent de $m = E(X)$.

1) $E(\overline{X}) = E(X) = m$ (l'estimateur est sans biais)

2) $V(\overline{X}) = V\left(\dfrac{\sum X_i}{n}\right) = \dfrac{1}{n^2}\sum V(X_i) = \dfrac{nV(X)}{n^2} = \dfrac{V(X)}{n}$

$V(\overline{X}) \to 0$ quand $n \to \infty$

Les conditions 1 et 2 du théorème étant vérifiées, l'estimateur $\overline{X}$ converge en probabilité vers $m = E(X)$, ce qui veut dire concrètement que si n est très grand $\overline{x}$ est très proche de $E(X)$.



densité de l'estimateur de m si n augmente $V(\overline{X})$ diminue

C. Estimateur usuel d'une proportion p

C'est la statistique usuelle, proportion observée dans un échantillon :

$$P_n = P_0 = \overline{X} = \frac{\sum X_i}{n} \text{ où } X_i \text{ suit une Bernoulli de paramètre } p.$$

Puisque $E(P_n) = p$ et $V(P_n) = \dfrac{pq}{n}$ tend vers zéro quand n tend vers l'infini, alors P_n converge en probabilité vers p.

Ce résultat veut dire que si la taille de l'échantillon est très grande, la proportion observée est presque certainement très proche de la valeur exacte p inconnue (voir l'application fiche n°17).

D. Estimateur usuel d'une moyenne $m = E(X)$

L'estimateur habituel est la statistique $\overline{X}$, moyenne de l'échantillon.

Puisque $E(\overline{X}) = m$ et $V(\overline{X}) = \dfrac{\sigma_X^2}{n}$, cet estimateur est sans biais et convergent. La moyenne $\overline{x}$ observée dans un échantillon se rapproche de $m = E(X)$ si n est très grand.

E. Estimateur usuel de la variance σ^2

• Si $m = E(X)$ est connu l'estimateur habituel est :

$$\boxed{\sum{}_n^2 = \frac{\sum_{i=1}^{n}(X_i - m)^2}{n}}$$

• Si $m = E(X)$ est inconnu l'estimateur habituel est :

$$\boxed{S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Après expériences, l'estimation ponctuelle de σ^2 est :

$$s_n^2 = s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Ce n'est pas tout à fait la variance de l'échantillon car on divise par $n - 1$ au lieu de n. Ces estimateurs ont été choisis parce qu'ils sont sans biais et convergents.

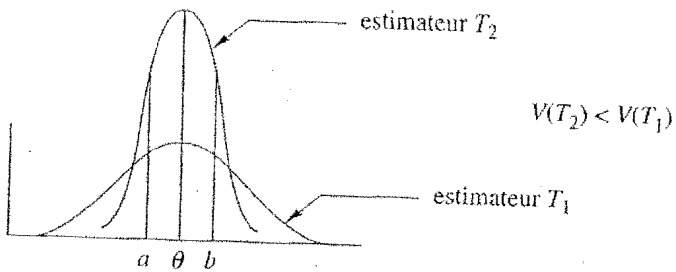
© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

III Compléments

A. Choix de l'estimateur dont la variance est la plus faible

Si la taille de l'échantillon est fixée et qu'il existe plusieurs estimateurs sans biais du paramètre θ, il est opportun de choisir celui dont la variance est la plus faible.



T_2 a une variance plus faible que T_1 et la probabilité qu'un intervalle (a, b) recouvre θ est supérieure si l'on choisit T_2 au lieu de T_1. Le critère de la plus petite variance n'est pas le seul critère pour choisir un estimateur, d'autres critères existent ils ne sont pas présentés ici.

B. Estimateur efficace

• Définition

Un estimateur sans biais est efficace s'il n'est pas possible de trouver un autre estimateur sans biais qui a une variance plus faible.

• Quantité d'information $I_n(\theta)$

$$I_n(\theta) = nE\left[(\ln f(X, \theta)'_\theta\right]^2$$

$I_n(\theta)$ est la quantité d'information apportée par un n-échantillon de la variable aléatoire X, f est la densité de probabilité de X.

• Inégalité de Cramer-Rao et efficacité

Sous des conditions très générales la variance d'un estimateur est toujours supérieure ou égale à $1/I_n(\theta)$, c'est l'inégalité de Cramer-Rao.

Si V (estimateur) $= 1/I_n(\theta)$, l'estimateur est efficace car il a la plus petite variance.

IV Application

ENONCE

On admet que le montant X des dépôts de chaque client, sur un compte bancaire, obéit à une loi normale de moyenne m et d'écart type σ ; m et σ sont inconnus et l'on propose de les estimer en tirant au hasard n clients parmi les titulaires de compte bancaire. Soit X_i la variable aléatoire : montant des dépôts du client n° i, i variant de 1 à n.

1. Donner les estimateurs habituels de m et σ^2 et montrer que l'estimateur habituel de m est sans biais et convergent en probabilité.

2. On a tiré au hasard 10 clients et obtenu les montants suivants des dépôts exprimés en francs :

5 700, 6 900, 13 500, 27 000, 10 000, 12 000, 14 000, 4 000, 8 600, 9 000.

Donner les estimations de m et σ associées à ces expériences.

SOLUTION

1. • Estimateur habituel de m : $\overline{X} = \dfrac{\sum X_i}{10}$, d'après le cours :

$E(\overline{X}) = m$ donc $\overline{X}$ est sans biais et puisque $V(\overline{X}) = \dfrac{\sigma^2}{n}$ tend vers zéro quand n tend vers l'infini alors $\overline{X}$ est convergent.

• Estimateur habituel de σ^2 lorsque m est inconnu,

$$S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

2. Les estimations correspondant aux réalisations x_i des variables X_i sont :

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_{10}}{10} = \frac{5\,700 + 6\,900 + \ldots + 9\,000}{10} = 11\,070 \text{ francs}$$

$$s = \sqrt{\sum(x_i - \overline{x})^2/9} = 6\,47,1 \text{ francs}$$

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

Fisher

21 Méthode du maximum de vraisemblance

I Objectifs

• Les estimations proposées dans la fiche précédente reposent sur l'idée intuitive qu'un moment inconnu d'une variable aléatoire ou d'une population peut être approché ou estimé par le moment équivalent calculé sur un échantillon. Ainsi, le moment d'ordre 1, c'est-à-dire l'espérance mathématique ou la moyenne dans une population est habituellement estimé par la moyenne des observations issues d'un échantillon. Cette méthode d'estimation par les moments ne convient pas dans toutes les applications pratiques et dans certains cas, elle ne fournit pas les meilleures estimations. C'est pour cette raison que le statisticien Ronald Fisher proposa, dans les années 1920, la méthode du maximum de vraisemblance.

• Cette méthode consiste à rechercher l'estimation du paramètre inconnu qui rend le plus probable ou le plus vraisemblable l'échantillon observé. Puisqu'il s'agit de trouver un maximum, cette méthode fait appel à la notion de dérivée en mathématiques, tout au moins pour les cas où la loi de probabilité est une fonction dérivable. Les estimateurs obtenus par la méthode du maximum de vraisemblance ont de bonnes propriétés statistiques. Certains estimateurs habituels obtenus par la méthode des moments sont les mêmes que ceux obtenus par la méthode du maximum de vraisemblance. La connaissance de cette méthode n'est pas indispensable pour traiter les intervalle de confiance (fiches n°22 à 24).

II L'essentiel à savoir

A. Les limites de la méthode des moments

Nous allons illustrer par un exemple classique le fait que la méthode des moments n'est pas toujours la plus pertinente.

La densité de probabilité f d'une variable X est $f(x) = 1/\theta$ si $0 \leqslant x \leqslant \theta$ et $f(x) = 0$ ailleurs.

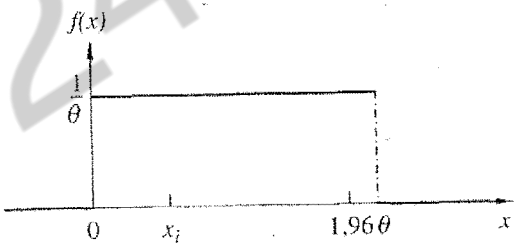
Cette loi étudiée dans la fiche n°14 est la loi uniforme, elle conduit à $E(X_i) = \dfrac{\theta}{2}$ et $E(\overline{X}) = \dfrac{\theta}{2}$. Si le paramètre θ est inconnu, son estimateur sans biais, déterminé par la méthode des moments à partir d'un n-échantillon de $X(X_1, X_2, \ldots, X_i, \ldots, X_n)$, est égal à $2\overline{X}$ puisque $E(2\overline{X}) = \theta$. Il est facile de montrer également que $2\overline{X}$ est un estimateur convergeant de θ. Après expériences, considérons la réalisation suivante d'un échantillon de taille 10 de X :

$$1{,}04 \ ; \ 0{,}95 \ ; \ 0{,}29 \ ; \ 0{,}12 \ ; \ 1{,}20 \ ; \ 1{,}13 \ ; \ 0{,}51 \ ; \ 0{,}89 \ ; \ 1{,}96 \ ; \ 1{,}36$$

Ces données numériques conduisent à l'estimation de θ suivante :

$$\widehat{\theta}_1 = 2\overline{x} = \frac{2\sum x_i}{10} = \frac{(1{,}04 + 0{,}95 + \ldots + 1{,}36)}{5} = 1{,}89$$

θ est estimé par $\widehat{\theta}_1 = 1{,}89$ par la méthode des moments.



Toutefois, puisque les réalisations de X sont comprises entre 0 et θ, il est évident que x_i est inférieur à θ, donc θ est supérieur à tous les x_i et en particulier au plus grand des x_i.

$$\theta \geqslant \operatorname*{Sup}_{1 \leqslant i \leqslant 10} x_i \ \operatorname{Sup} x_i = 1{,}96 \qquad \text{donc } \widehat{\theta} > 1{,}96$$

Ainsi la valeur $\widehat{\theta}_2 = 1{,}96$ est « plus vraisemblable » que $\widehat{\theta}_1 = 1{,}89$.

La plupart des estimateurs du maximum de vraisemblance ne sont pas aussi simples à déterminer et il faut utiliser des opérateurs de dérivation ou des méthodes numériques.

B. Fonction de vraisemblance

Soit S_X l'ensemble des réalisations d'une variable aléatoire X, discrète ou continue, qui dépend d'un paramètre θ.

• Si X est discrète, sa loi est définie par les quantités $P_\theta(x) = P_\theta(X = x)$.

La vraisemblance de la réalisation d'un n échantillon $(x_1, x_2, \ldots, x_i, \ldots, x_n)$ est la fonction L définie ci-après :

© Dunod – La photocopie non autorisée est un délit.

WWW.FALIB24.COM

$$L(x_1, x_2, \ldots, x_n, \theta) = P[(X_1 = x_1) \cap (X_2 = x_2) \cap \ldots \cap (X_n = x_n)]$$
$$= P(X_1 = x_1)P(X_2 = x_2) \cdot \ldots P(X_n = x_n)$$
$$= \prod_{i=1}^{n} P(X_i = x_i) \text{ ou } \prod_i P_\theta(X_i = x_i)$$

Le produit est possible dans la mesure où les variables X_i sont indépendantes.

EXEMPLE
Loi de Poisson : $L(X) = P(\lambda)$, $P(X = x_i) = e^{-\lambda}\dfrac{\lambda^{x_i}}{x_i!}$

$$L(x_1, \ldots, x_2, \ldots, x_n, \lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{\sum_i x_i}}{\prod_i x_i!}$$

• Si X est continue, sa loi est définie par sa densité $f_\theta(x)$. Par analogie avec le cas discret, la vraisemblance L s'écrit :

$$L(x_1, x_2, \ldots, x_n, \theta) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \ldots \cdot f_\theta(x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

EXEMPLE
Loi normale : $L(X) = N(m, \sigma)$, $f(x_i) = \dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x_i - m}{\sigma}\right)^2}$, cette loi dépend de m et σ :

$$L(x_1, x_2, \ldots, x_n, m, \sigma) = \prod_i \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x_i - m}{\sigma}\right)^2} = \frac{1}{2\pi^{n/2}\sigma^n}e^{-\frac{1}{2\sigma^2}\Sigma(x_i - m)^2}$$

C. Estimateur du maximum de vraisemblance

L'estimation du maximum de vraisemblance est la valeur $\widehat{\theta}$ qui, substituée à θ dans la fonction de vraisemblance $L(x_1, x_2, \ldots, x_n, \theta)$, rend celle-ci maximum, c'est-à-dire que la valeur observée du n-échantillon $(x_1, x_2, \ldots, x_n)$ est la plus probable pour $\theta = \widehat{\theta}$. Les valeurs de x_i étant fixées, L ne dépend plus que d'une seule variable θ, le maximum de la fonction, si elle est dérivable, correspond à :

$$L'_\theta = \frac{\partial L(x_1, x_2, \ldots, x_n, \theta)}{\partial \theta} = 0$$

La dérivée par rapport à θ est nulle.

La solution $\widehat{\theta}$ de cette équation correspond à un maximum si la dérivée seconde s'annule et change de signe en ce point. Pour la plupart des lois classiques, l'estimation du maximum de vraisemblance de $(x_1, x_2, \ldots, x_n)$ est la valeur de θ qui annule la dérivée première.

© Dunod – La photocopie non autorisée est un délit.

Remarque
- L'estimation du maximum de vraisemblance est : $\widehat{\theta} = T(x_1, x_2, \ldots, x_n)$
  L'estimateur du maximum de vraisemblance est : $\widehat{\Theta} = T(X_1, X_2, \ldots, X_n)$
- $\widehat{\theta}$ solution de $L'_\theta = 0$ est aussi solution de $(\ln L)'_\theta = 0$.
  La fonction ln désigne le logarithme néperien. Très souvent, on utilise les logarithmes car les calculs sont plus simples.

EXEMPLE
Pour la loi de Poisson de paramètre λ, la vraisemblance est :

$$L(x_1, x_2, \ldots, x_n, \lambda) = e^{-n\lambda}\frac{\lambda^{\Sigma x_i}}{\prod x_i!}$$

$$\ln L = -n\lambda + (\sum x_i)\ln\lambda - \ln\prod(x_i!)$$

$$(\ln L)'_\lambda = -n + \frac{\sum x_i}{\lambda} = 0$$

La solution de cette équation donne l'estimation du maximum de vraisemblance.

$$\widehat{\lambda} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \overline{x}$$

L'estimateur du maximum de vraisemblance est :

$$\widehat{\Lambda} = \frac{\sum X_i}{n} = \overline{X}$$

Cet estimateur est le même que celui obtenu par la méthode des moments pour $\lambda = E(X)$, car, le moment d'ordre 1 de l'échantillon est $\overline{X}$.

III Compléments

A. Cas où la loi de probabilité dépend de plusieurs paramètres

Une condition nécessaire pour qu'une fonction de plusieurs variables admette un extremum est que les dérivées partielles d'ordre 1 par rapport aux variables soient nulles.

EXEMPLE
La loi de X est une loi normale : $L(X) = N(m, \sigma)$
La vraisemblance est :

$$L(x, x, \ldots, x, m, \sigma) = \frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_i(x_i - m)^2}$$

Les conditions du 1er ordre $(\ln L)'_m = 0$ et $(\ln L)'_{\sigma^2} = 0$ conduisent à :

$$\widehat{m} = \frac{\sum x_i}{n} \text{ et } \widehat{\sigma^2} = \frac{\sum(x_i - \overline{x})^2}{n}$$

WWW.TALIB24.COM

Les estimateurs correspondants sont :

$$\widehat{M} = \frac{\sum X_i}{n} = \overline{X} \quad \text{et} \quad S'^2 = \frac{\sum (X_i - \overline{X})^2}{n}$$

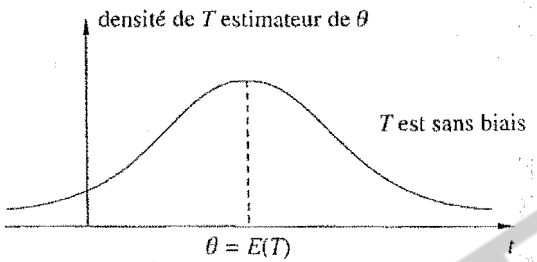
Ce sont les estimateurs du maximum de vraisemblance. Pour faire la démonstration mathématique correctement, il faut vérifier les conditions du second ordre. Ces estimateurs sont les mêmes que ceux obtenus par la méthode des moments.

B. Propriétés des estimateurs du maximum de vraisemblance

Les estimateurs du maximum de vraisemblance ont de bonnes propriétés statistiques. Soit T l'estimateur du maximum de vraisemblance d'un paramètre θ.

• Propriété 1 (biais)

T est sans biais ou asymptotiquement sans biais
$E(T) = \theta$ ou $E(T) \to \theta$ quand $n \to \infty$



• Propriété 2 (convergence)

T converge en probabilité vers θ : $T \xrightarrow[\text{proba}]{} \theta$

• Propriété 3 (efficacité)

La loi de probabilité de T tend vers une loi normale de moyenne θ et de variance $1/I_n(\theta)$.
$I_n(\theta)$ est la quantité d'information.
$V(T) \to 1/I_n(\theta)$ exprime que l'estimateur est asymptotiquement efficace.

IV Application

ÉNONCÉ

1. Donner l'estimation du maximum de vraisemblance pour une variable de Bernoulli.

2. Dans une population très nombreuse de possesseurs d'automobiles, la proportion p de personnes qui ont l'intention de racheter une voiture neuve dans les douze mois est inconnue.
Comment prélever un échantillon de 1 000 personnes et à quelle estimation de p conduisent ces observations si 185 personnes ont fait part de leur intention de racheter une voiture ?

3. L'estimation de p obtenue reflète-t-elle réellement une estimation précise de la proportion de personnes qui vont racheter une voiture dans les douze mois ?

SOLUTION

1. $L(X) = \mathcal{B}(1,p)$ d'où $P(X = x) = p^x(1-p)^{1-x}$, $x = 0$ ou 1

$$L(x_1, \ldots, x_2, \ldots, x_n, p) = \prod_i p^{x_i}(1-p)^{1-x_i} = p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$$

$$\ln L = (\sum x_i) \ln p + (n - \sum x_i) \ln (1-p)$$

La solution du maximum de vraisemblance conduit à $\widehat{p}$.

$$(\ln L)'_p = (\sum x_i)\frac{1}{p} - (n - \sum x_i) \cdot \frac{1}{1-p} = 0$$

Soit $\widehat{p} = \dfrac{\sum x_i}{n} = \overline{x}$ est l'estimation et : $\widehat{P} = \dfrac{\sum X_i}{n} = \overline{X}$ est l'estimateur.

La proportion observée dans l'échantillon est l'estimation du maximum de vraisemblance de p.

2. L'échantillon doit être prélevé au hasard et avec remise. Toutefois ici, comme la population est très nombreuse, le fait de faire des tirages sans remise exerce peu d'influence sur le résultat. Des méthodes comme celle des quotas donne de bonnes estimations, mais le calcul de marges d'erreur par les intervalles de confiance utilise des tirages au hasard. L'estimation du maximum de vraisemblance est :

$$\widehat{p} = p_0 = \frac{\sum x_i}{n} = \frac{185}{1\,000} = 0,185 = 18,5\ \%$$

3. L'incertitude due au fait que le renseignement est obtenu sur une partie de la population peut être mesurée par un intervalle de confiance (voir fiche n°23). Par contre, un autre type d'incertitude s'ajoute à la première, à savoir qu'il s'agit d'intentions d'achats et qu'il n'est pas possible de connaître sans des études complémentaires comment vont se concrétiser ces intentions d'achat dans les douze mois qui suivent.

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

22 Intervalles de confiance d'une moyenne

I Objectifs

• La valeur moyenne m d'une variable quantitative est inconnue dans une population. La façon la plus pertinente d'approcher cette valeur ou de l'estimer ponctuellement est de calculer la moyenne $\overline{x}$ sur un échantillon issu de cette population. Pour avoir une idée de la précision du résultat, il faut, à partir de $\overline{x}$, construire un intervalle qui a une forte probabilité de recouvrir m, c'est-à-dire un intervalle de confiance.

• Pour définir précisément cette notion, il est nécessaire de formaliser le problème avec les outils utilisés en théorie des probabilités. C'est-à-dire que X est une variable aléatoire pour laquelle $m = E(X)$ est inconnu, la moyenne désigne ici en fait l'espérance mathématique, un n-échantillon de X noté $X_1, X_2, \ldots, X_n$ conduit à définir un estimateur de m. Ensuite la loi de probabilité de l'estimateur étant connue, il est possible de déterminer les bornes de l'intervalle de confiance. De nombreuses notions définies dans cette fiche sont valables pour un paramètre θ quelconque et non pas seulement pour $\theta = m = E(X)$.

II L'essentiel à savoir

Il est rappelé qu'un n-échantillon de X est une suite de variables indépendantes et de même loi que X

A. Définition d'un intervalle de confiance

Par définition, un intervalle de confiance au niveau $1-\alpha$ pour m est un intervalle dont les extrémités sont aléatoires et qui recouvre m avec une probabilité $1-\alpha$. Soit $IC_{1-\alpha}(m)$ cet intervalle :

$$\boxed{P[IC_{1-\alpha}(m) \text{ recouvre } m] = 1-\alpha}$$

La probabilité que l'intervalle ne recouvre pas m est α. Les nombres α et $1-\alpha$ étant des probabilités sont compris entre 0 et 1.

$$IC_{1-\alpha}(m) = [A, B]$$

$$A = A(X_1, \ldots, X_n) \quad \text{et} \quad B = B(X_1, \ldots, X_n)$$

A et B sont les bornes de l'intervalle, ce sont des quantités qui dépendent du n échantillon $X_1, X_2, \ldots X_n$, ces quantités sont donc aléatoires.
Après n expériences, la suite $x_1, x_2, \ldots, x_n$ conduit aux bornes : $A(x_1, x_2, \ldots, x_n)$ et $B(x_1, x_2, \ldots, x_n)$. Ces deux valeurs numériques donnent une réalisation d'un intervalle de confiance. Souvent, dans les applications, lorsqu'il est demandé un intervalle de confiance, ce sont uniquement les valeurs numériques qu'il faut déterminer. La construction d'un intervalle de confiance avec des bornes aléatoires est plutôt une question de cours.

B. Intervalles à risques symétriques et unilatéraux

Soit $IC_{1-\alpha}(\theta)$ un intervalle de confiance au niveau $1-\alpha$ pour θ

$$P(A \leqslant \theta \leqslant B) = 1-\alpha \quad \text{et} \quad P(\theta < A) + P(\theta > B) = \alpha$$

En posant $\alpha_1 + \alpha_2 = \alpha$ et $P(\theta < A) = \alpha_1$ et $P(\theta > B) = \alpha_2$, on constate qu'il existe, tout au moins pour une variable aléatoire estimateur de θ continu, une infinité de façons de choisir α_1 et α_2, donc une infinité de choix pour A et B.

• Intervalles à risques symétriques

Le risque α est divisé par deux, soit $\alpha_1 = \alpha_2 = \alpha/2$

$$P(A > \theta) = P(B < \theta) = \alpha/2$$

L'intervalle $[A, B]$ est un intervalle bilatéral à risques symétriques. En l'absence de précisions dans une application, on détermine un intervalle à risques symétriques.

• Intervalles unilatéraux

Lorsque ($\alpha_1 = 0$ et $\alpha_2 = \alpha$) ou ($\alpha_1 = \alpha$ et $\alpha_2 = 0$), l'intervalle est unilatéral, car le risque est d'un seul côté de l'intervalle.

$$P(\theta > B) = \alpha \quad \text{et} \quad P(\theta < A) = 0$$

$$\text{ou} \quad P(\theta > B) = 0 \quad \text{et} \quad P(\theta < A) = \alpha$$

C. Intervalle de confiance d'une moyenne m avec écart type σ connu et pour la loi normale

$L(X) = N(m, \sigma)$, σ est connu.
Pour déterminer l'intervalle de confiance, on se propose d'utiliser $\overline{X}$ qui est un bon estimateur de m, en ce sens qu'il est sans biais, convergent et efficace.

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum X_i}{n}$$

© Dunod – La photocopie non autorisée est un délit.

WWW.FALIB24.COM

est une somme de variables normales indépendantes, sa loi de probabilité est donc bien connue :

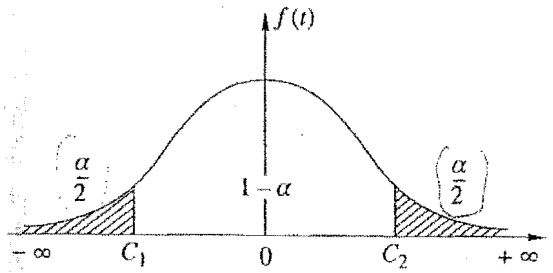
$$L(\overline{X}) = N\left(m, \frac{\sigma}{\sqrt{n}}\right) \text{ et si } T = \frac{\overline{X} - m}{\sigma/\sqrt{n}},\ L(T) = N(0,1)$$

De $P\left[C_1 \leqslant \dfrac{\overline{X} - m}{\sigma/\sqrt{n}} \leqslant C_2\right] = 1 - \alpha$

En prenant des risques symétriques il vient :

$$\alpha/2 = P\left[\frac{\overline{X} - m}{\sigma/\sqrt{n}} > C_2\right] = P[T > C_2] \text{ et } \alpha/2 = P\left[\frac{\overline{X} - m}{\sigma/\sqrt{n}} < C_1\right] = P[T < C_1]$$

On obtient C_1 et C_2 en utilisant les tables de la loi normale centrée réduite.



$$C_1 = -C_2 = t = t_{\text{table}}$$

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Ainsi $C_2 = -C_1 = t_{1-\alpha/2}$ car $P[T < t_{1-\alpha/2}] = 1 - \alpha/2$

On note aussi souvent $C_2 = -C_1 = t_{\alpha/2} = t_{\text{table}} = t$

Considérons l'événement E défini de la façon suivante :

$$E = \left[-t \leqslant \frac{\overline{X} - m}{\sigma/\sqrt{n}} < t\right] \Leftrightarrow \left[\overline{X} - t\frac{\sigma}{\sqrt{n}} \leqslant m < \overline{X} + t\frac{\sigma}{\sqrt{n}}\right]$$

et puisque $P[E] = 1 - \alpha$: $\left[\overline{X} - t\dfrac{\sigma}{\sqrt{n}},\ \overline{X} + t\dfrac{\sigma}{\sqrt{n}}\right]$ est un intervalle de confiance à risques symétriques de niveau $1 - \alpha$.

$$\boxed{IC_{1-\alpha}(m) = \left[\overline{X} - t\frac{\sigma}{\sqrt{n}}, \overline{X} + t\frac{\sigma}{\sqrt{n}}\right]}$$

Si $(x_1, x_2, \ldots, x_n)$ est une réalisation d'un n-échantillon de X, on obtient une réalisation de l'intervalle de confiance :

$$\boxed{IC_{1-\alpha}(m) = \left[\overline{x} - t\frac{\sigma}{\sqrt{n}}, \overline{x} + t\frac{\sigma}{\sqrt{n}}\right]}$$

EXEMPLE

Une variable X suit une loi normale de moyenne m inconnue et d'écart type $\sigma = 2$, donner la réalisation d'un intervalle de confiance au niveau $0,95$ pour m sachant que 20 expériences ont conduit à :

$$\sum_{i=1}^{20} x_i = 220$$

Il suffit d'appliquer directement le cours explicité ci-avant :

$$IC_{0,95}(m) = \left[\overline{x} \pm t_{0,975}\frac{\sigma}{\sqrt{n}}\right]$$

$$\sigma = 2,\quad \sqrt{n} = \sqrt{20},\quad \overline{x} = \frac{220}{20} = 11,\quad t_{0,975} = t_{\text{table}} = t = 1,96$$

$$IC_{0,95}(m) = \left[11 \pm 1,96\frac{2}{\sqrt{20}}\right] = [11 \pm 0,88] = [10,12\,;11,88]$$

Il y a 95 % de chances que la valeur exacte de m soit comprise entre $10,12$ et $11,88$.

Remarque

La longueur de l'intervalle est $\dfrac{1,96 \cdot 2}{\sqrt{n}}$ et si la taille n de l'échantillon est multiplié par 4, la longueur de l'intervalle est divisée par 2. Si n devient infini, la longueur de l'intervalle devient nulle quel que soit le niveau de confiance $1 - \alpha$. Ceci veut dire que $\overline{x}$ peut être rendu aussi proche de m que l'on veut. Ceci illustre la notion de convergence en probabilité de $\overline{X}$ vers m.

D. Intervalle de confiance d'une moyenne *m* avec écart type inconnu pour la loi normale

$L(X) = N(m, \sigma)$, m et σ sont inconnues ; c'est la situation que l'on rencontre le plus souvent dans la pratique.

L'intervalle précédent $\left[\overline{x} - t\dfrac{\sigma}{\sqrt{n}}, \overline{x} + t\dfrac{\sigma}{\sqrt{n}}\right]$ ne peut convenir dans la mesure où les bornes de l'intervalle dépendent d'un paramètre inconnu σ. Les statisticiens ont résolu ce problème en construisant la loi de Student qui est issue de la loi normale. Cette loi utilise l'écart type estimé s au lieu de l'écart type exact σ. La variable de Student est définie dans la fiche n°16.

$$T_{n-1} = \frac{\overline{X} - m}{S/\sqrt{n}} \text{ suit une loi de Student à } n - 1 \text{ degrés de libertés.}$$

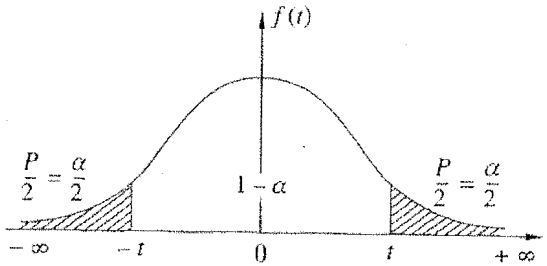
$$\overline{X} = \frac{\sum X_i}{n} \quad \text{et} \quad S^2 = \frac{\sum (X_i - \overline{X})^2}{n - 1}$$

Après expériences, $\overline{X}$ et S^2 deviennent $\overline{x}$ et s^2, T_{n-1} devient $t_{n-1} = \dfrac{\overline{x} - m}{s/\sqrt{n}}$

© Dunod – La photocopie non autorisée est un délit.

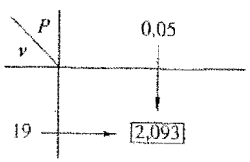
WWW.TALIB24.COM

De $P(-t < T_{n-1} < +t) = 1 - \alpha$, on déduit de la table de Student à $n-1$ degrés de libertés la valeur de t.



EXEMPLE

Si $n = 20$, $1 - \alpha = 95\,\%$, on lit dans la table de Student à l'intersection de la ligne « 19 degrés de libertés » et de la colonne $P = \alpha = 5\,\%$, $t = t_{\text{table}} = 2,093$ (pour la loi normale on obtient $t = 1,96$).



• Intervalle de confiance

Par analogie avec la loi normale, l'intervalle de confiance pour m à risques symétriques $\alpha/2$ est :

$$IC_{1-\alpha}(m) = \left[\overline{X} - t\frac{S}{\sqrt{n}}, \overline{X} + t\frac{S}{\sqrt{n}}\right]$$

Après expériences, la réalisation d'un intervalle est :

$$IC_{1-\alpha}(m) = \left[\overline{x} - t\frac{s}{\sqrt{n}}, \overline{x} + t\frac{s}{\sqrt{n}}\right]$$

Puisque $P(T_{n-1} < t) = 1 - \alpha/2$, il est plus correct de noter $t_{n-1}(1 - \alpha/2)$ au lieu de t. Cette valeur est aussi souvent notée $t_{\alpha/2}$ puisque $P(T > t_{\alpha/2}) = \alpha/2$.

III Compléments

A. Approximation par la loi normale

Les intervalles de confiance précédents utilisent la loi de $\overline{X}$ estimateur de m, la loi de $\overline{X}$ est une loi normale lorsque la variable X suit une loi normale. Lorsque X suit une autre loi que la loi normale, il est théoriquement possible de déterminer la loi de $\overline{X}$ et d'obtenir un intervalle de confiance par un procédé analogue à celui utilisé lorsque $\overline{X}$ suit une loi normale. Ceci peut conduire à des calculs ardus et ennuyeux. Là encore le théorème central limite constitue un auxiliaire précieux. L'utilisation de l'approximation par la loi normale à partir de $n = 30$ conduit au résultat suivant :

> $X_1, X_2, \ldots, X_i, \ldots, X_n$ est un n-échantillon de X, c'est-à-dire les X_i suivent la même loi et sont indépendantes.
> Si $n \geqslant 30$, $\overline{X}$ suit approximativement une loi normale. D'où la réalisation approchée de l'intervalle de confiance au niveau $1 - \alpha$
>
> $$IC_{1-\alpha}(m) = \left[\overline{x} \pm t\frac{\sigma_X}{\sqrt{n}}\right] \text{ ou } \left[\overline{x} \pm t\frac{s}{\sqrt{n}}\right]$$
>
> Avec : $P(-t < T < t) = 1 - \alpha$, $L(T) = N(0,1)$

Les résultats sur les intervalles de confiance obtenus avec la loi normale, restent approximativement valables avec d'autres lois, ceci traduit la notion de robustesse en statistique mathématique.

B. Loi de Student

$$T_{n-1} = \frac{N(0,1)}{\sqrt{\dfrac{\chi^2_{n-1}}{n-1}}} = \frac{\dfrac{\overline{X} - m}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\sum\left(\dfrac{X_i - \overline{X}}{\sigma}\right)^2}{n-1}}} = \frac{\overline{X} - m}{S/\sqrt{n}} \quad \text{car } S^2 = \frac{\sum(X_i - \overline{X})^2}{n-1}$$

Cette transformation a permis de faire disparaître le paramètre σ et de déterminer une intervalle de confiance pour la moyenne m quand l'écart type σ est inconnu.

IV Application

ENONCE

On admet que le montant X des avoirs de chaque client sur un compte bancaire suit une loi normale de moyenne m et d'écart type σ inconnu, m et σ sont inconnus. Ils sont estimés à partir du tirage au hasard de 10 clients ; les résultats en francs de ces tirages sont les suivants :

5 700, 6 900, 13 500, 27 000, 10 000, 12 000, 14 000, 4 000, 8 600, 9 000

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

1. Donner un intervalle de confiance à 95 % pour m.

2. Donner un intervalle de confiance à 90 % pour m.

3. Si les estimations de m et σ sont inchangées mais qu'elles portent sur 90 clients au lieu de 10, comment varient les résultats aux questions 1 et 2 ?

SOLUTION

1. $IC_{0,95}(m) = \left[\overline{x} \pm t_{n-1}\dfrac{s}{\sqrt{n}}\right]$

Les valeurs numériques conduisent à :

$$\overline{x} = \frac{\sum x_i}{10} = 11\,070 \text{ francs et } s^2 = \frac{\sum(x_i - \overline{x})^2}{10-1} = (647)^2$$

$$P(-t < T_{n-1} < +t) = 0,95 = 1 - \alpha$$

On lit dans la table de Student à 9 degrés de libertés $t_9 = 2,262$, d'où :

$$IC_{0,95}(m)\left[11\,070 \pm 2,262 \cdot \frac{647}{\sqrt{10}}\right] = [11\,070 \pm 463]$$

$$IC_{0,95}(m) = [10\,607;\ 11\,533]$$

Le montant moyen des avoirs estimé à 11 070 francs a 95 % de chances de se situer entre 10 607 francs et 11 533 francs.

2. Pour une probabilité $1 - \alpha = 0,90$, la valeur lue dans la table est $t_9 = 1,833$, d'où l'intervalle de confiance :

$$IC_{0,90}(m) = \left[11\,070 \pm 1,833 \cdot \frac{647}{\sqrt{10}}\right] = [10\,695,\ 11\,445]$$

3. $IC_{1-\alpha}(m) = \left[\overline{x} \pm t\dfrac{s}{\sqrt{n}}\right]$

La longueur de l'intervalle est $2t\dfrac{s}{\sqrt{n}}$, si $n = 90$ au lieu de 10, la longueur de l'intervalle est divisée par $\sqrt{9} = 3$.

Les calculs numériques sont immédiats.

$2t\dfrac{S}{\sqrt{n}}$

23 Intervalles de confiance d'une proportion

I Objectifs

• La proportion p d'individus qui présentent un caractère A dans une population est inconnue. La façon la plus pertinente d'approcher cette valeur ou de l'estimer ponctuellement est de calculer la proportion observée p_n sur un échantillon issu de cette population. Pour avoir une idée de la précision du résultat, il faut, à partir de p_n, construire un intervalle qui a une forte probabilité de recouvrir p. C'est un intervalle de confiance. La construction d'un tel intervalle repose sur la loi de probabilité d'un estimateur de cette proportion. L'estimateur classiquement utilisé est la variable aléatoire qui se déduit de la proportion observée p_n dans l'échantillon. Nous allons, à l'aide des résultats déjà connus, expliciter la loi de probabilité de cet estimateur et en déduire un intervalle de confiance, c'est-à-dire un intervalle qui recouvre p avec une probabilité $(1 - \alpha)$.

II L'essentiel à savoir

A. Loi de probabilité exacte de l'estimateur P_n de p

X est une variable de Bernoulli de paramètre p,
$X_i = 1$ si l'individu i présente le caractère A avec la probabilité p.
$X_i = 0$ si l'individu i ne présente pas le caractère A.
L'estimateur habituel est :

$$P_n = P_0 = \overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum X_i}{n}$$

Cet estimateur, qui peut être obtenu par la méthode des moments ou par le maximum de vraisemblance, est naturel dans la mesure où une estimation p_n associée à cet esti-

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

mateur est la proportion observée dans l'échantillon.

$$p_n = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}$$

L'estimateur P_n de p est sans biais, convergent, efficace.
La loi de P_n est connue, puisqu'au numérateur la quantité $\sum X_i$ représente une somme de Bernoulli indépendante de paramètre p, donc une binomiale de paramètre n et p.
La construction de l'intervalle de confiance à partir d'une loi binomiale serait bien ardue au niveau des calculs. Là encore l'approximation par la loi normale est d'une grande utilité pratique.

B. Approximation de la loi de l'estimateur P_n par la loi normale

Une somme de variables aléatoires indépendantes et de même loi peut être approximée sous certaines conditions par une loi normale.

Ainsi $P_n = \overline{X} = \frac{\sum X_i}{n}$ peut être approximée par une loi normale de paramètres $E(P_n) = p$ et $\sigma_{P_n} = \sqrt{\frac{pq}{n}}$ $\quad L(P_n)$ est approximativement $N\left(p, \sqrt{\frac{pq}{n}}\right)$

Les conditions les moins strictes pour faire cette approximation sont np et nq supérieurs à cinq (voir fiche n°18).

C. Intervalle de confiance de p

L'intervalle de confiance pour la moyenne m d'une variable aléatoire X qui suit une loi normale lorsque l'écart est connu est :

$$IC_{1-\alpha}(m)\left[\overline{X} - t\frac{\sigma_X}{\sqrt{n}},\ \overline{X} + t\frac{\sigma_X}{\sqrt{n}}\right]$$

où $L(T) = N(0,1)$ et $P(-t < T < +t) = 1 - \alpha$, $t = t_{1-\alpha/2}$
En transposant ce résultat à $P_n = \overline{X}$ il vient :

$$\left[P_n - t\sqrt{\frac{pq}{n}},\ P_n + t\sqrt{\frac{pq}{n}}\right]$$

Cet intervalle recouvre p avec la probabilité $1 - \alpha$, toutefois cet intervalle est inopérant et ce n'est pas un intervalle de confiance car les bornes dépendent du paramètre inconnu p. Dans la pratique il existe trois façons pour obtenir l'intervalle de confiance, nous donnons ici la plus classique qui consiste à remplacer le produit pq par $p_n\ q_n$, la réalisation d'un

intervalle de confiance au niveau $1 - \alpha$ pour p est ainsi :

$$IC_{1-\alpha}(p) = \left[p_n - t\sqrt{\frac{p_n q_n}{n}},\ p_n + t\sqrt{\frac{p_n q_n}{n}}\right]$$

Avec $P(-t < T < +t) = 1 - \alpha$ et $L(T) = N(0,1)$ si $(1 - \alpha) = 95\ \%$, $t = 1,96 \cong 2$, n est la taille de l'échantillon et $q_n = 1 - p_n$.

III Compléments

Les deux autres façons d'obtenir un intervalle de confiance pour p sont les suivantes :

A. Remplacement du produit pq par son maximum soit $\frac{1}{4}$

En remplaçant pq par son maximum soit $\frac{1}{4}$, on obtient un intervalle par excès, soit :

$$IC_{1-\alpha}(p)\left[p_n - t\sqrt{\frac{1}{4n}},\ p_n + t\sqrt{\frac{1}{4n}}\right]$$

B. Utilisation des abaques

Les abaques sont des courbes qui sont fournies par les tables statistiques usuelles. Elles permettent pour un niveau de confiance $1 - \alpha$ de 99 %, 95 % ou 90 % et pour une taille n d'échantillon d'obtenir directement les bornes de l'intervalle de confiance pour p. Cette méthode était très utilisée par les ingénieurs il y a quelques dizaines d'années pour les contrôles de fabrication en usine. À cette époque, les moyens de calcul modernes n'existaient pas ; ces abaques faisaient gagner un temps précieux pour les calculs d'intervalles de confiance à partir du pourcentage de pièces défectueuses dans un échantillon ou un lot issu de la fabrication.

EXEMPLE
La proportion observée $p_{obs} = p_n = 0,45$ est portée en abscisse. Si la taille de l'échantillon est $n = 100$, les bornes de l'intervalle de confiance se lisent en ordonnée par les intersections de la droite verticale $x = 45\ \%$ et la courbe qui correspond à $n = 100$.

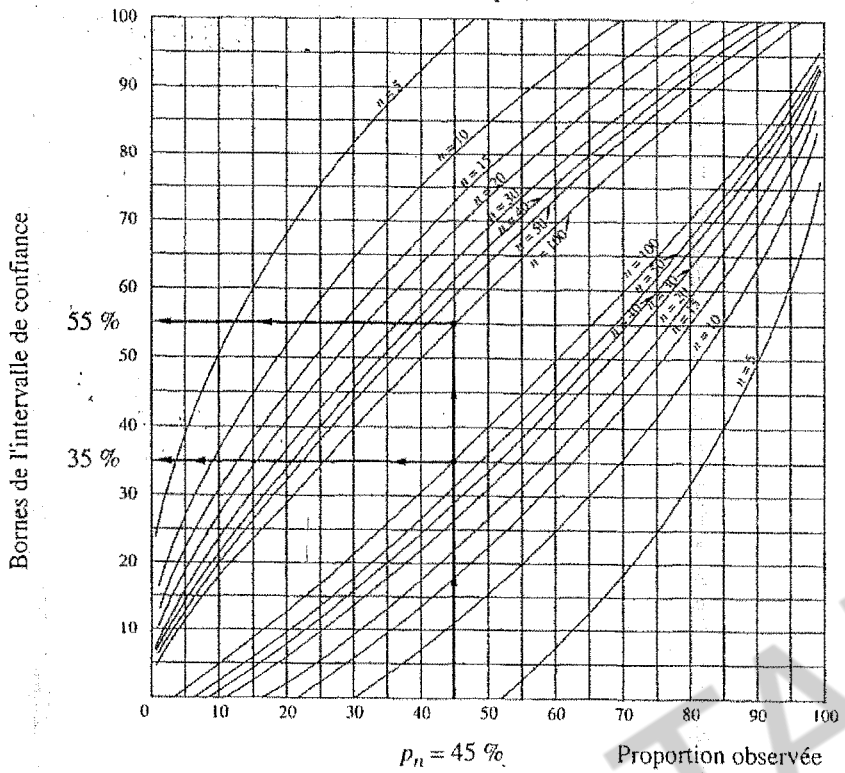
Donc, si $n = 100$, $p_n = 0,45$, $1 - \alpha = 95\ \%$

$$IC_{0,45}(p) = [0,35\,;0,55] = [35\ \%\,;45\ \%]$$

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

Intervalle de confiance à 95 % pour la proportion p de défectueux (abaque)



IV Application

ÉNONCÉ

Avant de lancer un nouvel article sur le marché, un industriel souhaiterait connaître la proportion p de personnes susceptibles d'acheter cet article. Pour cela, il fait réaliser une enquête auprès de 1 000 personnes.

1. Préciser comment il faut choisir les personnes qui vont répondre à l'enquête.

2. À la $i^{\text{ème}}$ personne enquêtée est associée la variable aléatoire X_i de Bernoulli de paramètre p.

a) Donner l'estimateur habituel de p et montrer qu'il est sans biais et convergent.

b) Sur 1000 personnes enquêtées, 77 souhaitent acheter cet article. Donner l'estimation de p correspondante.

3. a) Donner un intervalle de confiance au niveau 95 % pour p.

b) Donner la valeur de la taille de l'échantillon pour que la longueur de l'intervalle ne dépasse pas 2 %.

4. Donner un intervalle de confiance pour p au niveau 99 %.

SOLUTION

1. Les intervalles de confiance étudiés correspondent à des tirages au hasard d'observations qui sont les réalisations d'une même variable aléatoire X. Il n'est pas toujours possible de procéder ainsi dans la pratique, notamment lorsque la population est très dispersée géographiquement. On cherche alors à obtenir un échantillon représentatif vis-à-vis des critères susceptibles d'influer sur l'achat : l'âge, la catégorie sociale, le sexe, c'est la méthode des quotas.

2. a) Cette question est traitée dans cette fiche.

$X_i = 1$ avec la probabilité p s'il y a intention d'achat, et $X_i = 0$ sinon.

$$P_n = \overline{X} = \frac{\sum X_i}{n} \text{ est l'estimateur habituel de } p.$$

Puisque $E(P_n) = p$, P_n est sans biais.

Cette propriété, complétée par $V(P_n) = \dfrac{pq}{n}$ qui tend vers zéro quand n tend vers l'infini permet d'affirmer, d'après un théorème bien connu, que P_n converge en probabilité vers p.

b) L'estimation de p est la proportion p_n observée dans l'échantillon.

$$p_n = \overline{x} = \frac{\sum x_i}{n} = \frac{77}{1\,000} = 0,077 = 77\,\%$$

3.a)

$$IC_{0,95}(p) = \left[p_n \pm t_{\text{table}}\sqrt{\frac{p_n q_n}{n}}\right] = \left[0,077 \pm 1,96\frac{0,077(1-0,077)}{1\,000}\right]$$
$$= [0,0605\,;\,0,0935]$$

La proportion exacte est inconnue, mais elle a 95 % de chances d'être comprise entre 6,05 % et 9,35 %.

La réponse a été présentée directement. Pour un examen, il convient de préciser comment est obtenue la valeur 1,96 dans la table et de justifier que les données numériques permettent de faire l'approximation par la loi normale pour l'estimateur, c'est-à-dire qu'il faut vérifier que np et nq sont supérieurs à cinq aux bornes de l'intervalle.

© Dunod - La photocopie non autorisée est un délit.

WWW.TALIB24.COM

b) Longueur de l'intervalle : $l = 0,02$

$$l = 2 \times 1,96 \times \sqrt{\frac{0,077 \times 0,923}{n}} = 0,02$$

$$n = \left(\frac{2 \times 1,96}{0,02}\right)^2 \times 0,077 \times 0,923 = 2\,732 \quad (n \text{ entier})$$

c) $L(T) = N(0,1), \quad P(-t < T < +t) = 0,99$

d'où $$t = t_{\text{table}} = t_{0,995} = 2,576$$

$$IC_{0,99}(p) = \left[0,077 \pm 2,576\sqrt{\frac{0,077 \times 0,923}{1\,000}}\right] = [0,0553\,;0,0987]$$

24 Intervalles de confiance d'une variance

I Objectifs

La moyenne et la variance sont les paramètres les plus utilisés pour caractériser la distribution d'une variable quantitative X. Lorsque ces valeurs sont inconnues, les estimations ponctuelles fournies par un échantillon constituent des valeurs approchées de ces paramètres. Pour mesurer la précision des estimations, on construit des intervalles de confiance. Le principe est toujours le même, et à partir d'un bon estimateur de la variance – la loi de l'estimateur est connue, tout au moins de manière approchée – on détermine un intervalle qui recouvre la variance inconnue avec une probabilité $1 - \alpha$, cet intervalle est un intervalle de confiance pour la variance σ^2.

II L'essentiel à savoir

A. Les estimateurs classiques de la variance

$E(X) = m$, $V(X) = \sigma^2$, X_i sont des v.a. indépendantes de même loi que X, les estimateurs classiques de σ^2 sont :

$$\sum^2 = \sum_n^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{m})^2}{n} \quad \text{ou} \quad S_n^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

Le premier s'emploie lorsque m est connue, le second lorsque m est inconnue.
Après expériences $(x_1, x_2, \ldots, x_n)$ fournit les estimations suivantes de σ^2 :

$$\widehat{\sigma^2} = \frac{\sum(x_i - m)^2}{n} \quad \text{ou} \quad s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

Les estimateurs proposés sont sans biais et convergents.
$\sum^2$ est efficace (variance la plus faible). S^2 est asymptotiquement efficace.

© Dunod – La photocopie non autorisée est un délit.

WWW.TALB24.COM

B. Intervalle de confiance pour σ^2 lorsque m est connue

Nous choisissons pour X la loi normale, la loi de l'estimateur $\sum^2$ en dépend, nous allons la préciser.
$L(X) = N(m, \sigma)$ ainsi,

$$L\left(\frac{X_i - m}{\sigma}\right) = N(0,1) \quad \text{et} \quad \left(\frac{X_i - m}{\sigma}\right)^2 = [N(0,1)]^2 = \chi_1^2$$

$\displaystyle\sum_{i=1}^{n}\left(\frac{X_i - m}{\sigma}\right)^2$ suit donc une loi du chi-deux à n degrés de libertés (χ_n^2)

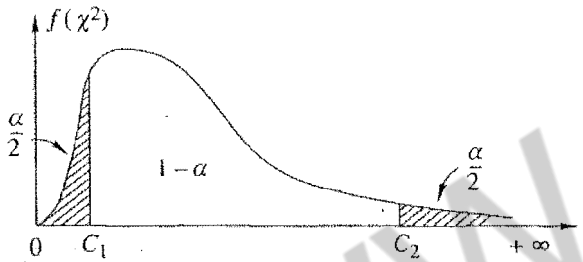
$$\frac{\sum(X_i - m)^2}{\sigma^2} = \frac{n}{\sigma^2}\sum\frac{(X_i - m)^2}{n} = n \cdot \frac{\sum_n^2}{\sigma^2} \quad \text{c'est-à-dire } L\left(\frac{n\sum_n^2}{\sigma^2}\right) = X_n^2.$$

La loi de l'estimateur est donc connue, et en se fixant un niveau de confiance $1 - \alpha$ de l'inégalité qui suit, il est possible de déduire un intervalle de confiance de niveau $1 - \alpha$ pour σ^2.

$$P\left(C_1 \leqslant \frac{n\sum_n^2}{\sigma} \leqslant C_2\right) = P(C_1 \leqslant \chi_n^2 \leqslant C_2) = 1 - \alpha$$

L'intervalle est généralement bilatéral et à risques symétriques, c'est-à-dire :

$$P(\chi_n^2 > C_2) = P(\chi_n^2 < C_1) = \frac{\alpha}{2}$$



$$C_1 = \chi_n^2(\alpha/2)$$
$$C_2 = \chi_n^2(1 - \alpha/2)$$

Considérons l'événement E défini de la façon suivante :

$$E = C_1 \leqslant n\frac{\sum_n^2}{\sigma^2} < C_2 \Leftrightarrow n\frac{\sum_n^2}{C_1} \leqslant \sigma^2 \leqslant n\frac{\sum_n^2}{C_2}$$

Nous avons ainsi obtenu un intervalle de confiance au niveau $1 - \alpha$ pour σ^2.

$$IC_{1-\alpha}(\sigma^2) = \left[n\frac{\sum_n^2}{C_2}\ ;\ n\frac{\sum_n^2}{C_1}\right] \text{ ou } \left[\frac{n\sum_n^2}{\chi_n^2(1-\alpha/2)}\ ;\ \frac{n\sum_n^2}{\chi_n^2(\alpha/2)}\right]$$

Après expériences :

$$IC_{1-\alpha}(\sigma)^2 = \left[\frac{n\widehat{\sigma}^2}{C_2}\ ;\ \frac{n\widehat{\sigma}^2}{C_1}\right]$$

EXEMPLE

$$L(X) = N(12, \sigma)\ ,\quad n = 20 \cdot \sum_{i=1}^{20} x_i^2 = 2\,960\ ,\ 1 - \alpha = 90\,\%$$

Le niveau de confiance est 90 %, ce qui donne des risques symétriques $\frac{\alpha}{2} = 5\,\%$.

Les valeurs C_1 et C_2 lues dans les tables d'une χ_{20}^2 sont : $C_1 = 10,85$ et $C_2 = 31,41$.

$$\widehat{\sigma}^2 = \frac{\sum(x_i - m)^2}{20} = \frac{\sum x_i^2}{20} - m^2 = \frac{2\,960}{20} - 144 = 4$$

La réalisation de l'intervalle de confiance est :

$$IC_{0,90}(\sigma^2) = \left[\frac{20 \times 4}{31,41}\ ;\ \frac{20 \times 4}{10,85}\right] = [2,55\ ;\ 7,37]$$

La variance est inconnue, l'échantillon observé aboutit à une estimation ponctuelle égale à 4, et la valeur exacte mais inconnue a 90 % de chances de se situer entre 2,55 et 7,37.

C. Intervalle de confiance pour σ^2 lorsque m est inconnue

$L(X) = N(m, \sigma)$, m et σ sont inconnues. $S^2 = \dfrac{\sum(X_i - \overline{X})^2}{n-1}$ est l'estimateur habituel de σ^2.

La loi de : $\dfrac{nS^2}{\sigma^2} = \displaystyle\sum_i\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 = \chi_{n-1}^2$

L'intervalle de confiance se construit comme au paragraphe précédent en utilisant S^2 au lieu de Σ^2 et la loi du chi-deux à $n - 1$ degrés de libertés au lieu de la loi du chi-deux à n degrés de libertés.

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^2}{C_2}\ ;\ \frac{(n-1)S^2}{C_1}\right]$$

Avec $P(\chi_{n-1}^2 > C_2) = P(\chi_{n-1}^2 < C_1) = \alpha/2$

Après expériences, la réalisation de l'intervalle de confiance est :

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s^2}{C_2}\ ;\ \frac{(n-1)s^2}{C_1}\right]$$

Dans la pratique, lorsque la variance est inconnue, la moyenne l'est aussi et l'intervalle de confiance qui vient d'être construit est celui que l'on rencontre le plus fréquemment.

© Dunod – La photocopie non autorisée est un délit.

WWW.TALIB24.COM

III Compléments

• Intervalle unilatéral

Les intervalles précédents de type $[A, B]$ sont bilatéraux et à risques symétriques car $P(\sigma^2 > B) = P(\sigma^2 < A) = \alpha/2$. (Ici les intervalles ne sont pas symétriques par rapport à σ^2, car la loi de probabilité ne l'est pas, contrairement à la loi normale utilisée pour l'intervalle de confiance d'une moyenne.)

Il est possible de déterminer un intervalle unilatéral en posant $P(\sigma^2 < A) = 0$ et $P(\sigma^2 > B) = \alpha$.

L'intervalle est unilatéral à gauche, bien sûr $A = 0$ car $\sigma^2 \geqslant 0$ et l'intervalle devient :

$$IC_{1-\alpha}(\sigma^2) = [0, B]$$

L'intervalle unilatéral à droite devient de la forme $[A, +\infty[$ et présente peu d'intérêt.

• Approximation de la loi du chi-deux par la loi normale

$$E(\chi^2_n) = n \qquad V(\chi^2_n) = 2n$$

Si $n \geqslant 30$, χ^2_n étant une somme de n variables χ^2_1 indépendantes peut être approximée par une loi normale.

$$L(\chi^2_n) \cong N(n, \sqrt{2n}) \quad \text{si} \quad n \geqslant 30$$

IV Application

ENONCE

La contenance d'une bouteille de lait est une variable aléatoire X qui suit une loi normale de moyenne m et d'écart type σ. Les paramètres m et σ sont inconnus. Pour estimer m et σ, il est prélevé un échantillon de taille $n = 25$. Soient $x_1, x_2, \ldots, x_i, \ldots, x_{25}$ les contenances respectives de 25 bouteilles prélevées au hasard dans la production. Ces prélèvements aboutissent à :

$$\sum_{i=1}^{25} x_i = 24,75 \text{ litres et } \sum_{i=1}^{25} (x_i)^2 = 24\,505\,956 \text{ litres} \times \text{litres}$$

1. Estimer m et σ^2.
2. Déterminer la réalisation d'un intervalle de confiance au niveau 0,95 pour m.
3. Déterminer un intervalle de confiance pour σ^2 au niveau de confiance 0,95.

© Dunod – La photocopie non autorisée est un délit.

SOLUTION

1. Estimation ponctuelle de m :

$$\widehat{m} = \overline{x} = \frac{\sum x_i}{25} = \frac{24,5}{25} = 0,99 \text{ litre}$$

Estimation ponctuelle de σ^2 :

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{25 - 1} = \frac{\sum (x_i)^2}{24} - \frac{25\overline{x}^2}{24} = 0,000\,144$$

$s = 0,012$ soit $1,2$ centilitres ou cm.

2. Intervalle de confiance pour la moyenne :

$$IC_{0,95}(m) = \left[\overline{x} \pm t\frac{s}{\sqrt{n}}\right]$$

$n = 25$, $n - 1 = 24$, $P(-t < T_{24} < +t) = 0,95$

T_{24} est la variable de Student à 24 degrés de libertés, d'où $t = 2,064$.

$$IC_{0,95}(m) = \left[0,99 \pm 2,064 \cdot \frac{0,012}{\sqrt{25}}\right] = [0,985\,;0,995]$$

Le volume moyen est inconnu, mais il a 95 % de chances d'être compris entre 0,985 et 0,995 litres.

3. Intervalle de confiance pour la variance :

$$IC_{0,95}(\sigma^2) = \left[\frac{25s^2}{C_2}, \frac{25s^2}{C_1}\right]$$

On choisit des risques symétriques : $P(\chi^2_{24} > C_2) = P(\chi^2_{24} < C_1) = 0,025$

$$P(\chi^2_{24} < C_2) = 0,975 \quad \text{d'où} \quad C_2 = \chi^2_{24}(0,975) = 39,37$$
$$P(\chi^2_{24} < C_1) = 0,025 \quad \text{d'où} \quad C_1 = \chi^2_{24}(0,025) = 12,40$$

D'où :

$$IC_{0,95}(\sigma^2) = \left[\frac{25 \times 0,000\,144}{39,37}\,;\ \frac{25 \times 0,000\,144}{12,40}\right] = [0,00\,009\,144\,;\ 0,0\,002\,932]$$

Les bornes sont exprimées en litres par litres. Il est plus judicieux de les exprimer en centilitres par centilitres. Il faut pour cela multiplier les chiffres précédents par $10^2 \times 10^2 = 10^4 = 10\,000$.

$$IC_{0,95}(\sigma^2) = [0,9144\,;\ 2,9032] \text{ centilitres par centilitres.}$$

Soit pour l'écart type σ :

$$IC_{0,95}(\sigma) = [0,95\,;\ 1,70] \text{ en centilitres.}$$

WWW.TALIB24.COM