Statistiques et probabilités appliquées
DUT - Techniques de Commercialisation

H. El-Otmany

ATER & docteur en mathématiques appliquées
Département Techniques de commercialisation
IUT de Tarbes
Email personnel : hamou.elotmany@gmail.com
Email professionnel : hammou.el-otmany@iut-tarbes.fr
Web site : www.hamoelotmany.github.io

Tarbes, 22 Novembre 2021

"Composition en caractère Cochin sur un ordinateur Apple MacBook Pro 15 à l'aide des logiciels libres suivants : LATEX 2_ε pdfTEX Xfig Grace

Motivations and applications

Motivations

- décrire les variables aléatoires sous la forme d'une expérience type ?
- analyser en détail pour pouvoir déduire les principales caractéristiques de toutes les expériences aléatoires.
- Décrire des phénomènes physiques, biologiques, économiques, ...

Thématique :

- Introduction aux probabilités
- Variables aléatoires, espérance et variance
- Lois de probabilités discrètes
- Lois de probabilités continues
- Échantillonnage et estimation
- Test χ^2 et l'indépendance des variables

Objectifs :

- calculer des probabilités sur les lois de probabilités
- associer des expériences aléatoires aux lois de probabilités
- utiliser les propriétés et la table de la loi normale pour effectuer des calculs de probabilités
- Calculer l'intervalle de confiance et tester l'indépendance en rapport avec des situations d'entreprises ;
- Savoir formuler une hypothèse et tester un risque : échantillonnage et estimation

Outline d'exposé

1. **Chapitre 1 : Lois de probabilités discrètes**
   - Introduction aux probabilités
   - Variables aléatoires, espérance et variance
   - Loi de Bernoulli et loi binomiale
   - Loi géométrique et loi hypergéométrique
   - Loi de Poisson ou modèle de Poisson
   - Approximation de la loi binomiale par la loi de Poisson
   - Ce qu'il faut retenir
2. **Chapitre 2 : Lois de probabilités continues**
   - Variables aléatoires continues
   - Loi uniforme
   - Loi normale ou de Laplace-Gauss
   - Approximation de la loi binomiale par la loi normale
   - Ce qu'il faut retenir
3. **Chapitre 3 : Échantillonnage et estimation**
   - Notions de la statistique inférentielle
   - Échantillonnage et estimation : définitions et applications
   - Échantillonnage : intervalle de fluctuation au risque d'erreur $1-\alpha$
   - Estimation et intervalle de confiance au risque d'erreur $1-\alpha$
   - Ce qu'il faut retenir
4. **Chapitre 4 : Test d'ajustement du khi-deux (χ^2)**
   - Introduction au test d'ajustement
   - Principe du test χ^2
   - Exemples d'application du χ^2
   - Ce qu'il faut retenir

Chapitre 1 : Lois de probabilités discrètes

1. Introduction aux probabilités
2. Variables aléatoires, espérance et variance
3. Loi de Bernoulli et loi binomiale
4. Loi géométrique et loi hypergéométrique
5. Loi de Poisson ou modèle de Poisson
6. Approximation de la loi binomiale par la loi de Poisson
7. Ce qu'il faut retenir

Variables aléatoires discrètes

Définitions

- Une variable aléatoire réelle (abbr. v.a.r) est une application **mesurable**

$$\begin{aligned} X : &\Omega \longrightarrow \mathbb{R} \\ &\omega \longmapsto X(\omega) \end{aligned}$$

- Une variable aléatoire est dite discrète si elle est définie sur un ensemble fini ou dénombrable.
- Soit X une variable aléatoire discrète, alors $0 \leq P(X = k) \leq 1$.

Exemples

- Nombre de "face ou pile" apparaissant après 10 jets d'une pièce.
- Nombre de véhicule passant à un carrefour dans une journée ;
- Lancé de dé ;
- Nombre de photons émis par une source lumineuse pendant une seconde ;
- Nombre de clients entrant dans un magasin le samedi.

Épreuve et loi de Bernoulli

Définition

Une épreuve de Bernoulli est une expérience aléatoire qui admet deux issues :

- le **succès S** ayant la probabilité p où $p \in]0;1[$;
- l'**échec E** ayant la probabilité $1-p$.

On définit la variable de Bernoulli par

$$X = \begin{cases} 1 & \texttt{si succès,} \\ 0 & \texttt{si échec.} \end{cases}$$

Exemples

- On lance un dé non truqué à 6 faces. On considère l'événement $S :=$" obtenir face de numéro 3".
- Une urne contient 5 boules rouges,3 boules vertes et 4 boules noires. On considère l'événement $S :=$"tirer une boule verte".

Remarque :

- La loi de probabilité de X est définie par $P(X=1) = P(S) = p$ et $P(X=0) = P(E) = 1-p$.
- On dit que la variable aléatoire X suit la loi de Bernoulli de paramètre p.
- Notation : $X \hookrightarrow \mathcal{B}(p)$.

Épreuve et loi de Bernoulli

Exemples

1. Succès dans un jeu :
   - Jeu de pile/face : $X = 1$ si face, $X = 0$ sinon. $\rightsquigarrow \mathcal{B}\left(\frac{1}{2}\right)$.
   - Lancé de dé : $X = 1$ si le résultat vaut 3, $X = 0$ sinon. $\rightsquigarrow \mathcal{B}\left(\frac{1}{6}\right)$.
2. Réponse oui/non dans un sondage :
   - Élections présidentielles 2022 : $X = 1$ si le président élu est M. Emmanuel Macron, $X = 0$ sinon. $\rightsquigarrow \mathcal{B}(p)$ avec p inconnu.
   - Vaccin COVID-19 : $X = 1$ si la personne approuve le vaccin, $X = 0$ sinon. $\rightsquigarrow \mathcal{B}(p)$, avec p inconnu.

Propriétés (paramètres de position et de dispersion)

Si $X \hookrightarrow \mathcal{B}(p)$ ($p \in]0;1[$). Alors, on a $E(X) = p$, $V(X) = p(1-p)$ et $\sigma_X = \sqrt{p(1-p)}$.

Exemple

- On lance un dé à 6 faces. Si on obtient le nombre 6, on a gagné, sinon on a perdu.
- Cette expérience est associée à une épreuve de Bernoulli. En effet, nous avons deux situations : succès (obtenir le nombre 6) et échec (ne pas obtenir le nombre 6).
- On note X la variable aléatoire prenant 0 (échec) et 1 (succès) $\rightsquigarrow X \hookrightarrow \mathcal{B}(\frac{1}{6})$.
- $E(X) = \frac{1}{6}$, $V(X) = \frac{1}{6}\left(1 - \frac{1}{6}\right) = \frac{5}{36}$, $\sigma_X = \frac{\sqrt{5}}{6}$.

Loi binomiale $\mathcal{B}(n;p)$

Définition

Soient n un entier supérieur ou égal à 2 et $p \in]0;1[$.

- On appelle schéma de Bernoulli de paramètres n et p, la répétition de n épreuves de Bernoulli indépendantes de paramètre p.
- On dit qu'une v.a.r X (nombre de succès obtenus) suit la loi binomiale de paramètres n et p lorsqu'elle comptabilise le nombre de succès d'un schéma de Bernoulli de paramètres n et p.
- Notation : $X \hookrightarrow \mathcal{B}(n;p)$.

Exemple

On reprend l'expérience aléatoire du lancé de dé. On la répète trois fois de manières indépendantes. À l'issue des trois lancers de dé, on compte le nombre total des succès.

- Cette expérience ayant deux situations : succès (obtenir le nombre 6) de probabilité $p = \frac{1}{6}$ et échec (ne pas obtenir le nombre 6) et $1 - p = \frac{5}{6}$.
- On répète cette expérience ($n = 3$) de manières indépendantes et identiques. Donc, c'est un schéma de Bernoulli.
- On note X la v.a.r associé au nombre de succès, donc $\rightsquigarrow X \hookrightarrow \mathcal{B}\left(3; \frac{1}{6}\right)$.

Remarque : soient $X_1, \cdots, X_n$ les v.a.r de Bernoulli respectives des n épreuves. Alors $X = X_1 + \cdots + X_n$ et $\mathcal{B}(p) = \mathcal{B}(1;p)$.

Loi binomiale $\mathcal{B}(n;p)$

Propriétés (paramètres de position et de dispersion)

Si $X \hookrightarrow \mathcal{B}(n;p)$ ($n \geq 2$ et $p \in]0;1[$), alors

- $E(X) = np$,
- $V(X) = np(1-p)$
- $\sigma_X = \sqrt{V(X)} = \sqrt{np(1-p)}$.

Exemple

On reprend l'expérience aléatoire du lancé de dé. On la répète trois fois de manières indépendantes. À l'issue des trois lancers de dé, on compte le nombre total des succès.

- On note X la v.a.r. associé au nombre de succès, donc $\rightsquigarrow X \hookrightarrow \mathcal{B}\left(3; \frac{1}{6}\right)$.
- $E(X) = 3 \times \frac{1}{6} = \frac{1}{2}$, $V(X) = 3 \times \frac{1}{6}\left(1 - \frac{1}{6}\right) = \frac{5}{12}$ et $\sigma_X = \sqrt{V(X)} = \sqrt{\frac{5}{12}}$.

Propriétés (somme de deux v.a. binomiales)

Si X et Y sont deux v.a. indépendantes telles que $X \hookrightarrow \mathcal{B}(n;p)$ et $Y \hookrightarrow \mathcal{B}(m;p)$, alors $X + Y \hookrightarrow \mathcal{B}(n+m;p)$.

Remarque : Deux v.a. X et Y définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$ sont indépendantes si les événements qu'elles engendrent sont indépendants.

Calcul de la loi de probabilité

Exemple

On reprend l'expérience aléatoire du lancé de dé. Si on obtient le nombre 6, on a gagné, sinon on a perdu. On la répète trois fois de manières indépendantes. À l'issue des trois lancers de dé, on compte le nombre total des succès.

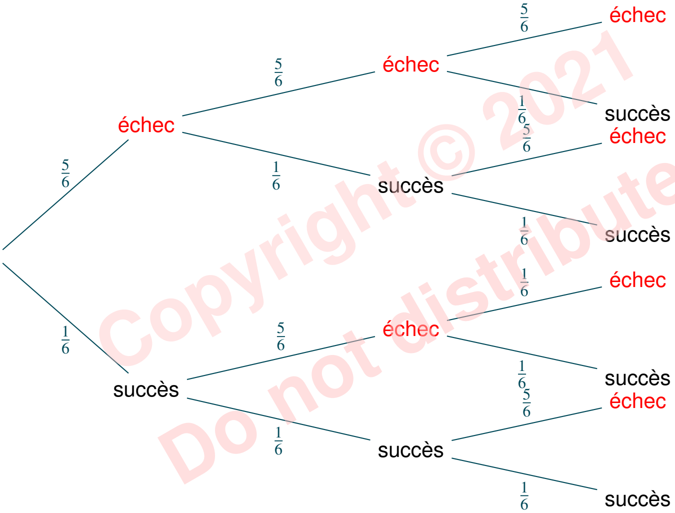
- On note X la v.a.r. associé au nombre de succès, donc $\rightsquigarrow X \hookrightarrow \mathcal{B}\left(3; \frac{1}{6}\right)$.
- On souhaite calculer la loi de probabilité de X.

Réponse : Commençons par remarquer que X peut prendre les valeurs $0, 1, 2$ et 3. En effet, lors des trois lancers, on peut avoir les situations suivantes :

- ne pas obtenir le nombre $'6'$, c'est-à-dire le nombre $'6'$ n'apparait pas lors des trois lancers (0 fois).
- obtenir une fois le nombre $'6'$.
- obtenir deux fois le nombre $'6'$.
- obtenir trois fois le nombre $'6'$.

Calcul de la loi de probabilité

Réalisons maintenant un arbre probabiliste qui traduit la situation où S représente le succès (obtenir le nombre $'6'$) de probabilité $\frac{1}{6}$, et E représente l'échec (ne pas obtenir le nombre $'6'$) de probabilité $1 - \frac{1}{6} = \frac{5}{6}$.



Calcul de la loi de probabilité

En analysant, l'arbre, on a :

- $P(X=0)$ désigne la probabilité d'obtenir 0 succès (i.e. obtenir trois échecs consécutifs). Il existe donc un seul chemin dans l'arbre contenant 0 succès et on a $P(X=0) = 1 \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = 1 \times \left(\frac{5}{6}\right)^3 = \frac{125}{216}$.
- $P(X=1)$ désigne la probabilité d'obtenir 1 succès. Il existe donc trois chemins dans l'arbre contenant 1 succès et on a $P(X=1) = 3 \times \left(\frac{1}{6}\right) \times \left(\frac{5}{6}\right)^2 = \frac{75}{216}$.
- $P(X=2)$ désigne la probabilité d'obtenir 2 succès. Il existe donc trois chemins dans l'arbre contenant 1 succès et on a $P(X=1) = 3 \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right) = \frac{15}{216}$.
- $P(X=3)$ désigne la probabilité d'obtenir 3 succès. Il existe donc un seul chemin dans l'arbre contenant 3 succès et on a $P(X=1) = 3 \times \left(\frac{1}{6}\right)^3 = \frac{1}{216}$.

D'où loi de probabilité de X (on vérifie facilement que la somme des probabilité vaut 1) :

Valeurs de x_i	0	1	2	3
Valeurs de $P(X=x_i)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Remarque : on vérifie facilement que la somme des probabilités vaut 1.

Calcul de la loi de probabilité

Théorème (loi de probabilité d'une v.a. binomiale)

Si X une v.a. suivant la loi Binomiale $B(n;p)$ ($n \geq 2$ et $p \in]0;1[$). Alors, pour tout $0 \leq k \leq n$, on a la loi de probabilité

$$P(X=k) = \underbrace{C_n^k}_{\text{nombre de chemins contenant k succès parmi n épreuves}} \times \underbrace{p^k}_{\text{probabilité des k succès}} \times \underbrace{(1-p)^{n-k}}_{\text{probabilité des n-k échecs}}$$

où $C_n^k = \begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$.

Remarque : Il est utile de noter que la probabilité de l'événement $\{X = k+1\}$ et celle de l'événement $\{X = k\}$ sont liées par :

$$P(X=k+1) = \frac{(n-k)p}{(k+1)(1-p)} P(X=k), k \in \{0, 1, \cdots, n-1\}.$$

Loi binomiale en fréquence ou en proportion

Description de l'expérience

- On considère X une variable aléatoire suivant la loi binomiale $\mathcal{B}(n;p)$ avec $n \in \mathbb{N}, n \geq 2$ et $0 \leq p \leq 1$ et associée au nombre de succès au cours de n épreuves.
- On note F_n la fréquence de succès définie par rapport entre le nombre de succès et le nombre d'épreuves.
- On écrit $F_n = \frac{X}{n}$.

Définition (loi binomiale en fréquence)

- On dit que la variable aléatoire F_n suit la loi binomiale en fréquence lorsqu'elle prend les valeurs d'une variable aléatoire suivant la loi binomiale divisées par le nombre d'épreuves.
- $P(F_n = f) = P(X = k) = C_n^k p^k (1-p)^{n-k}$ avec $k = n \times f$.
- Notation : $F_n \hookrightarrow BF(n;p)$

Exemple

Un procédé de fabrication produit 5% d'articles non conformes. Un échantillon de 50 unités de cet article est choisi (prélevé). Déterminer la probabilité pour qu'il y ait plus de 7% d'article non conformes dans l'échantillon.

Loi géométrique

Définition (loi géométrique)

- On dit que la variable aléatoire X suit la loi géométrique de paramètre p ($p \in]0;1[$) lorsqu'elle décrit la situation d'une répétition d'épreuves de Bernoulli indépendantes qu'il est nécessaire d'observer avant d'obtenir un succès.
- Loi de X : $P(X=k) = p(1-p)^{k-1}, \quad k \geq 1.$
- Notation : $X \hookrightarrow \mathcal{G}(p)$

Remarque : La loi géométrique est appelée aussi la loi "sans mémoire" (version discrète de la loi exponentielle). En effet, la connaissance du résultat des k premières expériences ne modifie pas les probabilités pour les suivantes.

Exemple : instant du premier succès dans un jeu de dé

- Lancer de dé à 6 faces ;
- X est le premier lancer pour lequel on obtient le nombre '6' ;
- $\rightsquigarrow X \hookrightarrow \mathcal{G}(\frac{1}{6})$;
- $P(X=5) = \frac{1}{6}\left(1-\frac{1}{6}\right)^{5-1} \approx 0.08037.$

Propriétés (paramètres de position et de dispersion)

Si $X \hookrightarrow \mathcal{G}(p)$, alors $E(X) = \frac{1}{p}$ et $V(X) = \frac{1-p}{p^2}$.

Loi géométrique

Propriété (absence de mémoire)

Si $X \hookrightarrow \mathcal{G}(p)$, alors pour tous $t, s > 0$:

$$P(X > s + t | X > t) = P_{X>t}(X > s + t) = P(X > s).$$

Exemple

On reprend l'expérience précédente : on lance un dé continuellement jusqu'à l'obtention du nombre '6'. Soit X le nombre de lancers nécessaires.

1. Calculer la moyenne, la variance, et l'écart-type de X.
2. Calculer la probabilité d'obtenir un premier '6' au 2^{me} lancer.
3. Calculer probabilité qu'il faille plus de 10 lancers pour obtenir un '6'.
4. Si aucun '6' n'a été obtenu lors des 8 premiers lancers, calculer probabilité qu'au moins deux autres lancers soient nécessaires.

Réponses : utiliser les propriétés des paramètres de position et de dispersion et celles de l'absence de mémoire.

Loi hypergéométrique

Description de l'expérience

Soit une urne contenant n boules dont p boules noirs et $n-p$ boules blanches. On effectue m tirages d'une boule sans remise avec $m \leq n$.

Définition

- Une variable aléatoire X suit la loi hypergéométrique de paramètres (m, n, p) si elle peut prendre les valeurs entières i telles que $0 \leq k \leq mp$; $n - mp \leq k \leq n$ et

$$P(X = k) = \frac{C_{np}^{k} C_{m-np}^{n-k}}{C_{m}^{n}}.$$

- $E(X) = np$, $V(X) = np(m-p)\frac{m-p}{m-1}$.
- Notation : $X \hookrightarrow \mathcal{H}(n; p; m)$.

Loi de Poisson ou modèle de Poisson

Définition

- Soit λ un réel strictement positif. La variable aléatoire X suit la loi de Poisson si elle est à valeurs dans l'ensemble des entiers (X prend les valeurs $0, 1, 2, \ldots$) et

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$

- $E(X) = \lambda$, $V(X) = \lambda$ et $\sigma_X = \sqrt{\lambda}$.
- $P(X = k) = \frac{\lambda}{k} P(X = k-1)$ pour $k \geq 1$.

Exemple

- Le nombre de clients qui fréquentent une banque chaque jours, pendant une heure suit la loi de Poisson de paramètre $\lambda = 7$.
- Calculer la probabilité qu'il y ait deux clients en heure. On a $P(X = 2) = \frac{7^2}{2!} e^{-7} \approx 0.0233$.
- Calculer la probabilité qu'il y ait au moins cinq clients fréquentant la banque en heure. On a : $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$, donc

$$P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - \left[\frac{7^0}{0!} e^{-7} + \frac{7^1}{1!} e^{-7} + \frac{7^2}{2!} e^{-7} + \frac{7^3}{3!} e^{-7} + \frac{7^4}{4!} e^{-7}\right] \approx 0.8271.$$

Loi de Poisson ou modèle de Poisson

Définition

- Soit λ un réel strictement positif. La variable aléatoire X suit la loi de Poisson si elle est à valeurs dans l'ensemble des entiers (X prend les valeurs $0, 1, 2, \ldots$) et

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots .$$

- $E(X) = \lambda$, $V(X) = \lambda$ et $\sigma_X = \sqrt{\lambda}$.
- $P(X = k) = \frac{\lambda}{k} P(X = k-1)$ pour $k \geq 1$.

Exemple

- Le nombre de clients qui fréquentent une banque chaque jours, pendant une heure suit la loi de Poisson de paramètre $\lambda = 7$.
- Calculer la probabilité qu'il y ait deux clients en heure. On a $P(X = 2) = \frac{7^2}{2!} e^{-7} \approx 0.0233$.
- Calculer la probabilité qu'il y ait au moins cinq clients fréquentant la banque en heure. On a : $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$, donc

$$P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - \left[\frac{7^0}{0!} e^{-7} + \frac{7^1}{1!} e^{-7} + \frac{7^2}{2!} e^{-7} + \frac{7^3}{3!} e^{-7} + \frac{7^4}{4!} e^{-7}\right] \approx 0.8271.$$

Loi de Poisson ou modèle de Poisson

Définition

- Soit λ un réel strictement positif. La variable aléatoire X suit la loi de Poisson si elle est à valeurs dans l'ensemble des entiers (X prend les valeurs $0, 1, 2, \ldots$) et

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$

- $E(X) = \lambda$, $V(X) = \lambda$ et $\sigma_X = \sqrt{\lambda}$.
- $P(X = k) = \frac{\lambda}{k} P(X = k - 1)$ pour $k \geq 1$.

Exemple

- Le nombre de clients qui fréquentent une banque chaque jours, pendant une heure suit la loi de Poisson de paramètre $\lambda = 7$.
- Calculer la probabilité qu'il y ait deux clients en heure. On a $P(X = 2) = \frac{7^2}{2!} e^{-7} \approx 0.0233$.
- Calculer la probabilité qu'il y ait au moins cinq clients fréquentant la banque en heure. On a : $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$, donc

$$P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - \left[\frac{7^0}{0!} e^{-7} + \frac{7^1}{1!} e^{-7} + \frac{7^2}{2!} e^{-7} + \frac{7^3}{3!} e^{-7} + \frac{7^4}{4!} e^{-7}\right] \approx 0.8271.$$

Loi de Poisson ou modèle de Poisson

Définition

- Soit λ un réel strictement positif. La variable aléatoire X suit la loi de Poisson si elle est à valeurs dans l'ensemble des entiers (X prend les valeurs $0, 1, 2, \ldots$) et

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$

- $E(X) = \lambda$, $V(X) = \lambda$ et $\sigma_X = \sqrt{\lambda}$.
- $P(X = k) = \frac{\lambda}{k} P(X = k - 1)$ pour $k \geq 1$.

Exemple

- Le nombre de clients qui fréquentent une banque chaque jours, pendant une heure suit la loi de Poisson de paramètre $\lambda = 7$.
- Calculer la probabilité qu'il y ait deux clients en heure. On a $P(X = 2) = \frac{7^2}{2!} e^{-7} \approx 0.0233$.
- Calculer la probabilité qu'il y ait au moins cinq clients fréquentant la banque en heure. On a : $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$, donc

$$P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$

$$= 1 - \left[\frac{7^0}{0!} e^{-7} + \frac{7^1}{1!} e^{-7} + \frac{7^2}{2!} e^{-7} + \frac{7^3}{3!} e^{-7} + \frac{7^4}{4!} e^{-7}\right] \approx 0.8271.$$

Loi de Poisson ou modèle de Poisson

Définition

- Soit λ un réel strictement positif. La variable aléatoire X suit la loi de Poisson si elle est à valeurs dans l'ensemble des entiers (X prend les valeurs $0, 1, 2, \ldots$) et

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, \ldots .$$

- $E(X) = \lambda$, $V(X) = \lambda$ et $\sigma_X = \sqrt{\lambda}$.
- $P(X = k) = \frac{\lambda}{k}P(X = k - 1)$ pour $k \geq 1$.

Exemple

- Le nombre de clients qui fréquentent une banque chaque jours, pendant une heure suit la loi de Poisson de paramètre $\lambda = 7$.
- Calculer la probabilité qu'il y ait deux clients en heure. On a $P(X = 2) = \frac{7^2}{2!}e^{-7} \approx 0.0233$.
- Calculer la probabilité qu'il y ait au moins cinq clients fréquentant la banque en heure. On a : $P(X \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4)$, donc

$$P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$$
$$= 1 - \left[\frac{7^0}{0!}e^{-7} + \frac{7^1}{1!}e^{-7} + \frac{7^2}{2!}e^{-7} + \frac{7^3}{3!}e^{-7} + \frac{7^4}{4!}e^{-7}\right] \approx 0.8271.$$

Calcul des probabilités via la table de loi de Poisson

Pour lire la valeur de $P(X = k)$ sur la table de Poisson, on procède comme suit :

- on détermine le paramètre de la loi de Poisson λ (i.e. $X \sim \mathcal{P}(\lambda)$)
- on se place sur la colonne λ .
- on se place sur la ligne k.
- on lit la valeur de $P(X = k)$ au point d'intersection de la colonne λ et la ligne k.

Exemple

On reprend l'exemple des clients qui fréquent une banque pendant une heure où $X \sim \mathcal{P}(7)$.

- Calculer $P(X = 2)$.
- On a $\lambda = 7$ et $k = 2$.

Extrait de la loi de Poisson

k \ λ	1	...	7	...
0	0.3679	...	0.0009	...
1	0.3679	...	0.0064	...
2	0.1839	...	$\boxed{0.0223}$	...
3	0.0613	...	0.0521	...
4	0.0153	...	0.0912	...
⋮	⋮	⋮	⋮	⋮

- Donc $P(X = 2) = \boxed{0.0223}$.
- Idem : $P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$
$= 1 - [0.0009 + 0.0064 + 0.0223 + 0.0521 + 0.0912] \approx 0.8271.$

Calcul des probabilités via la table de loi de Poisson

Pour lire la valeur de $P(X = k)$ sur la table de Poisson, on procède comme suit :

- on détermine le paramètre de la loi de Poisson λ (i.e. $X \sim \mathcal{P}(\lambda)$)
- on se place sur la colonne λ .
- on se place sur la ligne k.
- on lit la valeur de $P(X = k)$ au point d'intersection de la colonne λ et la ligne k.

Exemple

On reprend l'exemple des clients qui fréquent une banque pendant une heure où $X \sim \mathcal{P}(7)$.

- Calculer $P(X = 2)$.
- On a $\lambda = 7$ et $k = 2$.

Extrait de la loi de Poisson

k \ λ	1	...	7	...
0	0.3679	...	0.0009	...
1	0.3679	...	0.0064	...
2	0.1839	...	$\boxed{0.0223}$	...
3	0.0613	...	0.0521	...
4	0.0153	...	0.0912	...
⋮	⋮	⋮	⋮	⋮

- Donc $P(X = 2) = \boxed{0.0223}$.
- Idem : $P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$
  $= 1 - [0.0009 + 0.0064 + 0.0223 + 0.0521 + 0.0912] \approx 0.8271.$

Calcul des probabilités via la table de loi de Poisson

Pour lire la valeur de $P(X = k)$ sur la table de Poisson, on procède comme suit :

- on détermine le paramètre de la loi de Poisson λ (i.e. $X \sim \mathcal{P}(\lambda)$)
- on se place sur la colonne λ .
- on se place sur la ligne k.
- on lit la valeur de $P(X = k)$ au point d'intersection de la colonne λ et la ligne k.

Exemple

On reprend l'exemple des clients qui fréquent une banque pendant une heure où $X \sim \mathcal{P}(7)$.

- Calculer $P(X = 2)$.
- On a $\lambda = 7$ et $k = 2$.

Extrait de la loi de Poisson

k \ λ	1	...	7	...
0	0.3679	...	0.0009	...
1	0.3679	...	0.0064	...
2	0.1839	...	$\boxed{0.0223}$	...
3	0.0613	...	0.0521	...
4	0.0153	...	0.0912	...
⋮	⋮	⋮	⋮	⋮

- Donc $P(X = 2) = \boxed{0.0223}$.
- Idem : $P(X \geq 5) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)]$
  $= 1 - [0.0009 + 0.0064 + 0.0223 + 0.0521 + 0.0912] \approx 0.8271.$

Approximation de la loi Binomiale $\mathcal{B}(n;p)$ par la loi de Poisson $\mathcal{P}(\lambda)$

Théorème (utile en pratique)

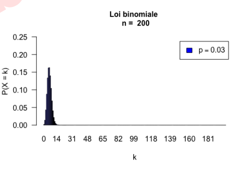
Soit $X \hookrightarrow \mathcal{B}(n;p)$ avec $n \geq 2$ et $p \in [0,1]$. Supposons que n et p vérifient :

- $n \geq 30$;
- $0 \leq p \leq 0.1$;
- $np < 10$.

Alors, on peut estimer la loi de $X \hookrightarrow \mathcal{B}(n;p)$ par celle de $Y \hookrightarrow \mathcal{P}(n \times p)$. De plus, on a $P(X=k) \approx P(Y=k) = \frac{(n \times p)^k}{k!} e^{-n \times p}$.

Exemple

- Probabilité de gagner une partie : 3%.
- Nombre de réalisations indépendantes : 200 parties
- X la variable aléatoire associée au nombre de parties gagnées.
- Sachant que $X \hookrightarrow \mathcal{B}(200;0.03)$, calculer $P(X=10)$?



- On a $n = 200 \geq 30$, $p = 0.03 \leq 0.1$ et $np = 200 \times 0.3 = 6 < 10$.
- D'après le théorème, on a $Y \hookrightarrow \mathcal{P}(6)$ et $P(X=10) \approx P(Y=10) = \frac{6^{10}}{10!} e^{-6} \approx 0.0413$

Approximation de la loi Binomiale $\mathcal{B}(n;p)$ par la loi de Poisson $\mathcal{P}(\lambda)$

Théorème (utile en pratique)

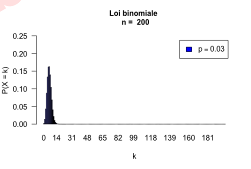
Soit $X \hookrightarrow \mathcal{B}(n;p)$ avec $n \geq 2$ et $p \in [0,1]$. Supposons que n et p vérifient :

- $n \geq 30$;
- $0 \leq p \leq 0.1$;
- $np < 10$.

Alors, on peut estimer la loi de $X \hookrightarrow \mathcal{B}(n;p)$ par celle de $Y \hookrightarrow \mathcal{P}(n \times p)$. De plus, on a $P(X = k) \approx P(Y = k) = \frac{(n \times p)^k}{k!} e^{-n \times p}$.

Exemple

- Probabilité de gagner une partie : 3%.
- Nombre de réalisations indépendantes : 200 parties
- X la variable aléatoire associée au nombre de parties gagnées.
- Sachant que $X \hookrightarrow \mathcal{B}(200; 0.03)$, calculer $P(X = 10)$?



  - On a $n = 200 \geq 30$, $p = 0.03 \leq 0.1$ et $np = 200 \times 0.3 = 6 < 10$.
  - D'après le théorème, on a $Y \hookrightarrow \mathcal{P}(6)$ et $P(X = 10) \approx P(Y = 10) = \frac{6^{10}}{10!} e^{-6} \approx 0.0413$

TD n°2 sur Moodle

Exercice 2 : Un agent immobilier a estimé que la probabilité de vendre un appartement suite à une visite était 7%. Il effectue en général 120 visites par mois.
On considère que les visites d'appartements sont des expériences aléatoires indépendantes les unes des autres. On appelle A la variable aléatoire égale au nombre d'appartement vendus en un mois après une visite.

1. Justifier que la variable aléatoire A suit une loi binomiale dont on précisera les paramètres n et p.
2. On souhaite calculer la probabilité que l'agent vende exactement 30 appartements en un mois après une visite.
   a) Calculer C_{120}^{10} à l'aide de la calculatrice. Que peut-on conclure .
   b) Justifier que l'on peut approcher la loi de A par une loi normale que l'on précisera.
   c) À l'aide de cette approximation, calculer la probabilité de vendre exactement 30 appartements en un mois après une visite.

Réponses :

1. Il s'agit d'une expérience aléatoire admettant deux issues :
   - succès : "vendre un appartement suite à une visite" de probabilité $p = 0.07$.
   - échec : "ne pas vendre un appartement suite à une visite" de probabilité $1 - p = 1 - 0.07 = 0.93$.

   Cette expérience aléatoire est une épreuve de Bernoulli. On répète cette expérience 120 fois de manière indépendante. Donc, c'est un schéma de Bernoulli.
   On note A la variable aléatoire associé au nombre d'appartement vendus après une visite en un mois, c'est-à-dire au nombre de succès obtenus à l'issue des 120 épreuves de Bernoulli. Par conséquent A suit la loi binomiale $\mathcal{B}(n = 120; p = 0.07)$.

TD n°2 sur Moodle

1. On souhaite calculer la probabilité que l'agent vende dix appartements en un mois après une visite.

   a) Calculer C_{120}^{10} à l'aide de la calculatrice.
   À l'aide de la calculatrice, one gets :

   $$C_{120}^{10} = 1.1606817863878 \times 10^{14}.$$

   Ce nombre est difficile à obtenir car il fait intervenir des factorielles importantes et la calculatrice nous donne qu'une valeur approchée (en réalité $C_{120}^{10} = 116068178638776$). De même, pour les valeurs 0.07^{10} et $(1-0.07)^{110}$ qui interviennent dans le calcul de $P(A=10)$. Pour cela, nous proposons d'approcher la loi de A par une loi convenable.

   b) Justifier que l'on peut approcher la loi de A par une loi de Poisson que l'on précisera.
   Comme $n = 120 \geq 30$, $p = 0.07 \leq 0.1$ et $np = 120 \times 0.07 = 8.4 < 10$, on peut donc approcher la loi de A par une loi de Poisson $\mathcal{P}(8.4)$.

   c) À l'aide de cette approximation, calculer la probabilité voulue.
   On a $P(X=10) = e^{-8.4} \times \frac{8.4^{10}}{10!} \approx 0.1084$.

TD n°2 sur Moodle

1. On souhaite calculer la probabilité que l'agent vende dix appartements en un mois après une visite.

   a) Calculer C_{120}^{10} à l'aide de la calculatrice.
   À l'aide de la calculatrice, one gets :

   $$C_{120}^{10} = 1.1606817863878 \times 10^{14}.$$

   Ce nombre est difficile à obtenir car il fait intervenir des factorielles importantes et la calculatrice nous donne qu'une valeur approchée (en réalité $C_{120}^{10} = 116068178638776$). De même, pour les valeurs 0.07^{10} et $(1-0.07)^{110}$ qui interviennent dans le calcul de $P(A=10)$. Pour cela, nous proposons d'approcher la loi de A par une loi convenable.

   b) Justifier que l'on peut approcher la loi de A par une loi de Poisson que l'on précisera.
   Comme $n = 120 \geq 30$, $p = 0.07 \leq 0.1$ et $np = 120 \times 0.07 = 8.4 < 10$, on peut donc approcher la loi de A par une loi de Poisson $\mathcal{P}(8.4)$.

   c) À l'aide de cette approximation, calculer la probabilité voulue.
   On a $P(X=10) = e^{-8.4} \times \frac{8.4^{10}}{10!} \approx 0.1084$.

TD n°2 sur Moodle

1. On souhaite calculer la probabilité que l'agent vende dix appartements en un mois après une visite.

   a) Calculer C_{120}^{10} à l'aide de la calculatrice.
   À l'aide de la calculatrice, one gets :

   $$C_{120}^{10} = 1.1606817863878 \times 10^{14}.$$

   Ce nombre est difficile à obtenir car il fait intervenir des factorielles importantes et la calculatrice nous donne qu'une valeur approchée (en réalité $C_{120}^{10} = 116068178638776$). De même, pour les valeurs 0.07^{10} et $(1-0.07)^{110}$ qui interviennent dans le calcul de $P(A=10)$. Pour cela, nous proposons d'approcher la loi de A par une loi convenable.

   b) Justifier que l'on peut approcher la loi de A par une loi de Poisson que l'on précisera.
   Comme $n = 120 \geq 30$, $p = 0.07 \leq 0.1$ et $np = 120 \times 0.07 = 8.4 < 10$, on peut donc approcher la loi de A par une loi de Poisson $\mathcal{P}(8.4)$.

   c) À l'aide de cette approximation, calculer la probabilité voulue.
   On a $P(X=10) = e^{-8.4} \times \frac{8.4^{10}}{10!} \approx 0.1084$.

TD n°2 sur Moodle

1. On souhaite calculer la probabilité que l'agent vende dix appartements en un mois après une visite.

   a) Calculer C_{120}^{10} à l'aide de la calculatrice.
   À l'aide de la calculatrice, one gets :

   $$C_{120}^{10} = 1.1606817863878 \times 10^{14}.$$

   Ce nombre est difficile à obtenir car il fait intervenir des factorielles importantes et la calculatrice nous donne qu'une valeur approchée (en réalité $C_{120}^{10} = 116068178638776$). De même, pour les valeurs 0.07^{10} et $(1-0.07)^{110}$ qui interviennent dans le calcul de $P(A=10)$. Pour cela, nous proposons d'approcher la loi de A par une loi convenable.

   b) Justifier que l'on peut approcher la loi de A par une loi de Poisson que l'on précisera.
   Comme $n = 120 \geq 30$, $p = 0.07 \leq 0.1$ et $np = 120 \times 0.07 = 8.4 < 10$, on peut donc approcher la loi de A par une loi de Poisson $\mathcal{P}(8.4)$.

   c) À l'aide de cette approximation, calculer la probabilité voulue.
   On a $P(X=10) = e^{-8.4} \times \frac{8.4^{10}}{10!} \approx 0.1084$.

TD n°2 sur Moodle

1. On souhaite calculer la probabilité que l'agent vende dix appartements en un mois après une visite.

   a) Calculer C_{120}^{10} à l'aide de la calculatrice.
   À l'aide de la calculatrice, one gets :

   $$C_{120}^{10} = 1.1606817863878 \times 10^{14}.$$

   Ce nombre est difficile à obtenir car il fait intervenir des factorielles importantes et la calculatrice nous donne qu'une valeur approchée (en réalité $C_{120}^{10} = 116068178638776$). De même, pour les valeurs 0.07^{10} et $(1-0.07)^{110}$ qui interviennent dans le calcul de $P(A = 10)$. Pour cela, nous proposons d'approcher la loi de A par une loi convenable.

   b) Justifier que l'on peut approcher la loi de A par une loi de Poisson que l'on précisera.
   Comme $n = 120 \geq 30$, $p = 0.07 \leq 0.1$ et $np = 120 \times 0.07 = 8.4 < 10$, on peut donc approcher la loi de A par une loi de Poisson $\mathcal{P}(8.4)$.

   c) À l'aide de cette approximation, calculer la probabilité voulue.
   On a $P(X = 10) = e^{-8.4} \times \frac{8.4^{10}}{10!} \approx 0.1084$.

Résumé de la loi de Poisson

Essentiel à retenir

1. La loi de Poisson permet la modélisation des événements (probabilité faible) où le futur est indépendant du passé.
2. $X \hookrightarrow \mathcal{P}(\lambda)$ avec $\lambda > 0$, on a $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.
3. $E(X) = \lambda$, $V(X) = \lambda$, $\sigma_X = \sqrt{\lambda}$
4. Approximation de la loi binomiale par celle de Poisson : soit $X \hookrightarrow \mathcal{B}(n; p)$. Si
   - $n \geq 30$;
   - $0 \leq p \leq 0.1$;
   - $np < 10$.

$\rightsquigarrow$ on peut estimer la loi de X par celle de Poisson telle que $Y \hookrightarrow \mathcal{P}(np)$ et

$$P(X = k) \approx P(Y = k) = \frac{(np)^k}{k!} e^{-np}.$$

Chapitre 2 : Lois de probabilités continues

Variables aléatoires continues

Définitions

- Une variable aléatoire réelle (abbr. v.a.r) est une application **mesurable**

$$\begin{aligned} X :& \Omega \longrightarrow \mathbb{R} \\ & \omega \longmapsto X(\omega) \end{aligned}$$

- Une variable aléatoire est dite continue si elle est définie sur un ensemble indénombrable. Autrement dit, elle prend les valeurs sur un intervalle ou sur tout $\mathbb{R}$.
- Soit X une variable aléatoire continue, alors $P(X = k) = 0$ pour tout $k \in \mathbb{N}$.

Exemples

- intervalle de temps entre 2 passages de train $X(\Omega) = [0, T]$;
- temps d'attente à la banque $X(\Omega) = \mathbb{R}^+$;
- longueur de cheveux $X(\Omega) = \mathbb{R}^+$;
- durée de vie en secondes d ?une pièce électronique $X(\Omega) = [0, T]$;
- poids à la naissance $X(\Omega) = [0, m]$

Variables aléatoires continues

Définitions

- Une variable aléatoire réelle (abbr. v.a.r) est une application **mesurable**

$$\begin{aligned} X :& \Omega \longrightarrow \mathbb{R} \\ & \omega \longmapsto X(\omega) \end{aligned}$$

- Une variable aléatoire est dite continue si elle est définie sur un ensemble indénombrable. Autrement dit, elle prend les valeurs sur un intervalle ou sur tout $\mathbb{R}$.
- Soit X une variable aléatoire continue, alors $P(X = k) = 0$ pour tout $k \in \mathbb{N}$.

Exemples

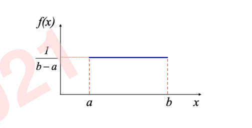
- intervalle de temps entre 2 passages de train $X(\Omega) = [0, T]$;
- temps d'attente à la banque $X(\Omega) = \mathbb{R}^+$;
- longueur de cheveux $X(\Omega) = \mathbb{R}^+$;
- durée de vie en secondes d ?une pièce électronique $X(\Omega) = [0, T]$;
- poids à la naissance $X(\Omega) = [0, m]$

Loi uniforme

Définition

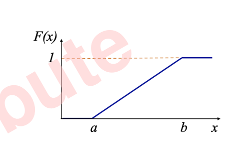
La loi de probabilité uniforme est une loi de probabilité absolument continue dont la fonction densité est proportionnelle à la longueur d'intervalle $[a, b]$ et définie par

$$x \longmapsto \frac{1}{b-a}\mathbb{I}_{a\leq x\leq b} = \begin{cases} \frac{1}{b-a} & si\ a \leq x \leq b \\ 0 & ailleurs \end{cases}$$



et sa fonction de répartition est

$$x \longmapsto \begin{cases} 0 & si\, x < a \\ \frac{x-a}{b-a} & si\ a \leq x \leq b \\ 1 & si\, x \geq b \end{cases}$$



Notation : $X \hookrightarrow \mathcal{U}(a, b)$.

Propriété

Soit $X \hookrightarrow \mathcal{U}([a;b])$, alors : $E(X) = \frac{b+a}{2}$, $\quad V(X) = \frac{(b-a)^2}{12}$, $\sigma_X = \sqrt{\frac{(b-a)^2}{12}}$.

Loi uniforme

Calculs des probabilités

Soient $X \hookrightarrow \mathcal{U}([a;b])$ et $[c;d] \subset [a;b]$. Alors, on a

$$P(c \leq X \leq d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}.$$

Remarque : Dans l'exemple précédent, comme dans toute variable aléatoire absolument continue, on a $P(X=0)=0$. En effet, pour tout réel a, on a : $\int_a^a f(x)dx = 0$.

Exemple

Dans une ville idéale, les autobus passent à chaque arrêt exactement toutes les $20 minutes$. On appelle X le temps d'attente en minutes d'un autobus à un arrêt. X est une variable aléatoire suivant la loi uniforme sur l'intervalle $[0,20]$, on a donc

$$P(5 \leq X \leq 15) = \frac{15-5}{20-0} = \frac{10}{20} = \frac{1}{2}, \quad P(X \geq 15) = P(15 \leq X \leq 20) = \frac{20-15}{15-0} = \frac{5}{15} = \frac{1}{3}.$$

Enfin, le temps d'attente moyen qui est égal à $E(X)$ vaut $\frac{0+20}{2}$, soit 10 minutes.

Loi normale centrée réduite ou loi gaussienne $\mathcal{N}(0;1)$

Définition

- La loi de probabilité normale s'applique aux phénomènes aléatoires ayant de nombreuses causes dont les effets s'ajoutent sans que l'un d'eux ne soit dominant.
- La loi de probabilité normale centrée réduite est une loi de probabilité continue dont la fonction densité est définie sur $\mathbb{R}$ par

$$x \longmapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Notation : $X \hookrightarrow \mathcal{N}(0;1)$.

Exemple : étude du nombre de clients qui fréquentent une banque.

On suppose que le nombre des clients suit une loi normale. En effet, il dépend de plusieurs facteurs explicatifs : emplacement de la banque, communication, prix proposé, produits (type de cartes, livret A, LDD, ...),facilité de paiement, ...

Loi normale centrée réduite ou loi gaussienne $\mathcal{N}(0;1)$

Définition

- La loi de probabilité normale s'applique aux phénomènes aléatoires ayant de nombreuses causes dont les effets s'ajoutent sans que l'un d'eux ne soit dominant.
- La loi de probabilité normale centrée réduite est une loi de probabilité continue dont la fonction densité est définie sur $\mathbb{R}$ par

$$x \longmapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Notation : $X \hookrightarrow \mathcal{N}(0;1)$.

Exemple : étude du nombre de clients qui fréquentent une banque.

On suppose que le nombre des clients suit une loi normale. En effet, il dépend de plusieurs facteurs explicatifs : emplacement de la banque, communication, prix proposé, produits (type de cartes, livret A, LDD, ...),facilité de paiement, ...

Loi normale centrée réduite ou loi gaussienne $\mathcal{N}(0; 1)$

Propriété

Soit X une variable aléatoire suivant une loi normale centrée réduite $X \hookrightarrow \mathcal{N}(0; 1)$, alors

$$E(X) = 0, \quad V(X) = 1.$$

Remarque : la fonction de répartition associée à $\mathcal{N}(0; 1)$ n'a pas de formule simple. Elle implémentée dans certains logiciels de calcul numérique sous le nom de **normcdf**(Matlab) ou **pnorm** (R) :

$$\Phi : x \longmapsto \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Exemple : étude de la capacité de mémoire d'adultes atteints d'une maladie neurologique

On suppose que chaque individu lit 30 mots et doit ensuite en réciter le plus possible.

- Population $\mathcal{P} = \{$patients atteints de la maladie$\}$.
- Variable quantitative $X := \{$nombre de mots retenus"$\}$.
- 2 paramètres μ et σ.

Loi normale $\mathcal{N}(\mu;\sigma^2)$

Définition

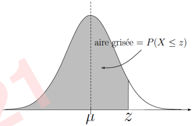
- Soit μ et σ deux réels, tels que $\sigma \neq 0$. On dit qu'une variable aléatoire X suit une loi normale de moyenne μ et variance σ^2 si la variable aléaoire
$$Z = \frac{X-\mu}{\sigma}$$
suit une loi normale centrée réduite $\mathcal{N}(0;1)$.
- Notation : $X \hookrightarrow \mathcal{N}(\mu;\sigma^2)$.
- La densité de X est la fonction $f : x \longmapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  $\rightsquigarrow$ cette formule n'est pas utile pour ce cours mais à connaître
- $X \hookrightarrow \mathcal{N}(\mu;\sigma^2)$, alors $E(X) = \mu$ et $V(X) = \sigma^2$.
- μ est le paramètre de position.
- σ est le paramètre de dispersion.

Attention : dans certains livres, il est noté $X \hookrightarrow \mathcal{N}(\mu;\sigma)$ au lieu de $X \hookrightarrow \mathcal{N}(\mu;\sigma^2)$.

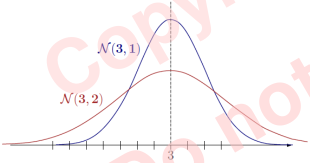
Caractéristiques de la loi normale $\mathcal{N}(\mu;\sigma^2)$

1 Caractéristiques de la loi normale $\mathcal{N}(\mu;\sigma^2)$

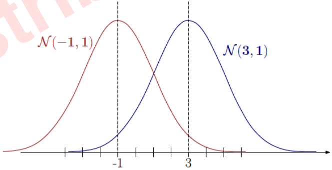
- courbe continue (tracé sans lever la main).
- courbe symétrique par rapport à la droite verticale $X = \mu$.
- forme en cloche.
- aire totale sous la courbe vaut 1 (donc l'aire allant de moins l'infini à 0 vaut 0.5).
- aire grisée représente la proportion cumulée.
- probabilité correspondante à la surface sous la courbe.



2 Exemples des lois normales



(a) $\mu_1 = \mu_2$ et $\sigma_1 \neq \sigma_2$ (left)



(b) $\mu_1 \neq \mu_2$ et $\sigma_1 = \sigma_2$ (right)

Formules de calcul des probabilités

1. Pour faire des calculs avec $X \hookrightarrow \mathcal{N}(0;1)$, on utilise sa table et

   - a est un nombre positif :
     - $P(X \leq a) = \Phi(a)$.
     - $P(X \geq a) = 1 - \Phi(a)$.
     - $P(X \leq -a) = P(X \geq a) = 1 - \Phi(a)$.
     - $P(X \geq -a) = \Phi(a)$.
     - $P(|X| \leq a) = P(-a \leq X \leq a) = 2 \times \Phi(a) - 1$.

   - a et b deux réels avec $a < b$: $P(a \leq X \leq b) = \Phi(b) - \Phi(a)$.

2. Pour faire des calculs avec $Y \hookrightarrow \mathcal{N}(\mu;\sigma^2)$ ($a > 0$), on procède comme suit :

   - on **centre** et on **réduit** Y pour se ramener à une variable suivant $\mathcal{N}(0;1)$
   - on pose donc $X = \frac{Y-\mu}{\sigma}$.
   - on calcule par exemple : $P(Y \leq a) = P\left(\frac{Y-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(X \leq \frac{a-\mu}{\sigma}\right)$.
   - on cherche la valeur de $\frac{a-m}{\sigma}$ dans la table de $\mathcal{N}(0;1)$.

Remarque : si $Y \hookrightarrow \mathcal{N}(m;\sigma^2)$ et $a > 0$, alors :

$$P(-a \leq Y \leq a) = P\left(\frac{-a-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(\frac{-a-\mu}{\sigma} \leq X \leq \frac{a-\mu}{\sigma}\right).$$

Calcul des probabilités via la table de loi normale $\mathcal{N}(0;1)$

Soit $X \hookrightarrow \mathcal{N}(0;1)$. Pour lire la valeur de $P(X \leq z)$ sur la table de la loi normale centrée réduite $\mathcal{N}(0,1)$, on procède comme suit :

- on précise la formule à utiliser : ici $P(X \leq z) = \Phi(z)$.
- on se place sur ligne $i.j$ où i est la 1^{re} unité et j est la 1^{re} décimale de z.
- on se place sur la colonne $0.0k$ telle que $i.j + 0.0k = z$ où k est la 2^{me} décimale de z.
- on lit la valeur de $\Phi(z)$ au point d'intersection de la ligne i, j et la colonne $0, 0k$.

Exemple

On suppose qu'une certaine variable $X \hookrightarrow N(0;1)$. Pour quelle proportion d'individus est-ce que $X \leq 1.44$?

- Calculer $P(X \leq 1.44) = \Phi(1.44)$.
- Chercher $z = 1.44$ dans la table.
- $i.j = 1.4$ et $0.0k$=0.04.

Extrait de la loi normale $\mathcal{N}(0,1)$

z	0.00	0.01	...	0.04	...
0.0	0.5000	0.5040	...	0.5160	...
0.1	0.5398	0.5438	...	0.5557	...
⋮	⋮	⋮	⋮	⋮	⋮
1.4	0.9192	0.9207	...	$\boxed{0.9236}$	...
⋮	⋮	⋮	⋮	⋮	⋮

- Donc $P(X \leq 1.44) = \boxed{0.9236}$.
- Idem : $P(X \leq -1.44) = P(X \geq 1.44) = 1 - P(X \leq 1.44) = 1 - \Phi(1.44) = 1 - 0.9236$

Calcul des probabilités via la table de loi normale $\mathcal{N}(0;1)$

Soit $X \hookrightarrow \mathcal{N}(0;1)$. Pour lire la valeur de $P(X \leq z)$ sur la table de la loi normale centrée réduite $\mathcal{N}(0,1)$, on procède comme suit :

- on précise la formule à utiliser : ici $P(X \leq z) = \Phi(z)$.
- on se place sur ligne $i.j$ où i est la 1^{re} unité et j est la 1^{re} décimale de z.
- on se place sur la colonne $0.0k$ telle que $i.j + 0.0k = z$ où k est la 2^{me} décimale de z.
- on lit la valeur de $\Phi(z)$ au point d'intersection de la ligne i,j et la colonne $0,0k$.

Exemple

On suppose qu'une certaine variable $X \hookrightarrow N(0;1)$. Pour quelle proportion d'individus est-ce que $X \leq 1.44$?

- Calculer $P(X \leq 1.44) = \Phi(1.44)$.
- Chercher $z = 1.44$ dans la table.
- $i.j = 1.4$ et $0.0k$=0.04.
- Donc $P(X \leq 1.44) = \boxed{0.9236}$.
- Idem : $P(X \leq -1.44) = P(X \geq 1.44) = 1 - P(X \leq 1.44) = 1 - \Phi(1.44) = 1 - 0.9236$

Extrait de la loi normale $\mathcal{N}(0,1)$

z	0.00	0.01	...	0.04	...
0.0	0.5000	0.5040	...	0.5160	...
0.1	0.5398	0.5438	...	0.5557	...
⋮	⋮	⋮	⋮	⋮	⋮
1.4	0.9192	0.9207	...	$\boxed{0.9236}$	...
⋮	⋮	⋮	⋮	⋮	⋮

Calcul des probabilités via la table de loi normale $\mathcal{N}(0;1)$

Soit $X \hookrightarrow \mathcal{N}(0;1)$. Pour lire la valeur de $P(X \leq z)$ sur la table de la loi normale centrée réduite $\mathcal{N}(0,1)$, on procède comme suit :

- on précise la formule à utiliser : ici $P(X \leq z) = \Phi(z)$.
- on se place sur ligne $i.j$ où i est la 1^{re} unité et j est la 1^{re} décimale de z.
- on se place sur la colonne $0.0k$ telle que $i.j + 0.0k = z$ où k est la 2^{me} décimale de z.
- on lit la valeur de $\Phi(z)$ au point d'intersection de la ligne i,j et la colonne $0,0k$.

Exemple

On suppose qu'une certaine variable $X \hookrightarrow N(0;1)$. Pour quelle proportion d'individus est-ce que $X \leq 1.44$?

- Calculer $P(X \leq 1.44) = \Phi(1.44)$.
- Chercher $z = 1.44$ dans la table.
- $i.j = 1.4$ et $0.0k$=0.04.
- Donc $P(X \leq 1.44) = \boxed{0.9236}$.

Extrait de la loi normale $\mathcal{N}(0,1)$

z	0.00	0.01	...	0.04	...
0.0	0.5000	0.5040	...	0.5160	...
0.1	0.5398	0.5438	...	0.5557	...
⋮	⋮	⋮	⋮	⋮	⋮
1.4	0.9192	0.9207	...	$\boxed{0.9236}$	...
⋮	⋮	⋮	⋮	⋮	⋮

- Idem : $P(X \leq -1.44) = P(X \geq 1.44) = 1 - P(X \leq 1.44) = 1 - \Phi(1.44) = 1 - 0.9236$

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

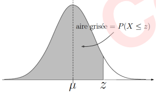
Table de $\mathcal{N}(0;1)$

0.0k

i.j

z	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	0.7703	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	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.99[illegible]
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.99[illegible]

$P(X \leq x)$ **avec**
$X \hookrightarrow \mathcal{N}(0;1)$



Feuille de Travaux dirigés

Exercice 4 : Soient $X \hookrightarrow \mathcal{N}(0;1)$ et $Y \hookrightarrow \mathcal{N}(5;2^2)$. Calculer les probabilités suivantes :

1. $P(X \leq 2) = \Phi(2) \approx 0.9772$
2. $P(X < -2.02) = 1 - \Phi(2.02) \approx 0.0217$
3. $P(X > 2.2) = 1 - \Phi(2.2) \approx 0.0139$
4. $P(X \geq -2.22) = \Phi(2.22) \approx 0.9868$
5. $P(-1.45 \leq X < 1.45) = 2\Phi(1.45) - 1 \approx 0.8530$
6. $P(0.57 < X \leq 1.82) = \Phi(1.82) - \Phi(0.57) \approx 0.2499$
7. $P(Y < 3.2) = P\left(\frac{Y-5}{2} < \frac{3.2-5}{2}\right) = P(X < -0.9) = 1 - \Phi(0.9) \approx 0.1841,$
8. $P(Y \geq 7.88) = P\left(\frac{Y-5}{2} \geq \frac{7.88-5}{2}\right) = P(X \geq 1.44) = 1 - \Phi(1.44) \approx 0.0749,$
9. $P(Y < 7.5) = P\left(\frac{Y-5}{2} < \frac{7.5-5}{2}\right) = P(X < 1.25) = \Phi(1.25) \approx 0.8944$
10. $P(3.96 \leq Y \leq 6.02) = P\left(\frac{3.96-5}{2} \leq \frac{Y-5}{2} \leq \frac{6.02-5}{2}\right) = P(-0.52 \leq X \leq 0.51) = P(Y \leq 0.51) - P(Y \leq -0.52) = \Phi(0.51) - (1 - \Phi(0.52)) \approx 0.3935$
11. $P(3.44 < Y < 6.78) = P\left(\frac{3.44-5}{2} < \frac{Y-5}{2} < \frac{6.78-5}{2}\right) = P(-0.78 < X < 0.89) = P(X < 0.89) - P(X < -0.78) = \Phi(0.89) - (1 - \Phi(0.78)) \approx 0.5956.$

Feuille de Travaux dirigés

Exercice 6 : En utilisant la modélisation statistique multilinéaire, un économiste français prédit que le prix du gasoil en mars 2022 suivra $\mathcal{N}(1.6; 2.44^2)$. Soit X la variable aléatoire donnant le prix du gasoil.

1. Calculer la probabilité pour que le prix du gasoil soit moins de 1.42€.

On cherche $P(X \leq 1.42)$:

$$P(X \leq 1.42) = P\left(\frac{X - 1.6}{2.44} \leq \frac{1.42 - 1.6}{2.44}\right) = P(Z \leq -0.071)$$

où $Z = \frac{X-1.6}{2.44} \hookrightarrow \mathcal{N}(0; 1)$.. Or, d'après la table de la loi normale

$$P(Z \leq -0.071) \approx P(Z \geq 0.07) = 1 - P(Z \leq 0.07) \approx 1 - 0.5279 \approx 0.4721.$$

Feuille de Travaux dirigés

Exercice 6 (suite) :

1. Calculer la limite α telle que la probabilité d'avoir un prix plus petit est de 40%.

   On veut calculer $P(X \leq \alpha) = 40\% = 0.4$. Pour ce faire,

   - on centre et on réduit X :
   $$P\left(X \leq \alpha\right) = P\left(\frac{X-1.6}{2.44} \leq \frac{\alpha-1.6}{2.44}\right) = P\left(Z \leq \frac{\alpha-1.6}{2.44}\right) = 0.2$$
   où $Z = \dfrac{X-1.6}{2.44}$ suit $\mathcal{N}(0;1)$.
   - On détermine α tel que $P\left(Z \leq \dfrac{\alpha-1.6}{2.44}\right) = 0.4$.
   - On compare la probabilité à 0.5 ; $0.2 \leq 0.5 \Longrightarrow \dfrac{\alpha-1.6}{2.44} \leq 0$.
   - On cherche $P\left(Z \leq -\frac{\alpha-1.6}{2.44}\right) = 1 - 0.4 = 0.6$.
   - On lit la table **à l'envers** : $0.26 \approx -\frac{\alpha-1.6}{2.44}$.Donc $0.26 \approx -\frac{\alpha-1.6}{2.44}$.
   On peut prendre $\rightsquigarrow \alpha \approx 1$€.

Approximation de la loi Binomiale $\mathcal{B}(n;p)$ par la loi normale $\mathcal{N}(\mu;\sigma^2)$

Théorème (utile en pratique)

Soit Z une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n;p)$. Supposons que n et p vérifient les trois conditions :

- $n \geq 20$;
- $np \geq 10$;
- $n(1-p) \geq 10$.

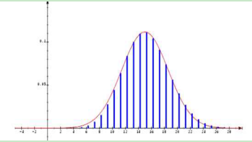
Alors on peut approcher la loi de Z par la loi de X suivant $\mathcal{N}(\mu;\sigma^2)$ avec $\mu = n \times p$ et $\sigma = \sqrt{n \times p \times (1-p)}$.

Exemple

Soit $Z \sim \mathcal{B}(200;0.35)$. Alors, on a :

- $n = 200 \geq 20$;
- $np = 200 \times 0.35 = 70 \geq 10$;
- $n(1-p) = 200 \times (1-0.35) = 130 \geq 10$.

Donc, on peut estimer la loi de Z par la loi de X, où $X \hookrightarrow \mathcal{N}(\mu;\sigma^2)$ avec $\mu = np = 70$ et $\sigma = \sqrt{np(1-p)} = \sqrt{45.5}$.



$\rightsquigarrow$ En bleu, $\mathcal{B}(200;0.35)$ et en rouge $\mathcal{N}(70;45.5)$.

TD n°2 sur Moodle

Exercice 8 : Un agent immobilier a estimé que la probabilité de vendre un appartement suite à une visite était 15%. Il effectue en général 120 visites par mois.
On considère que les visites d'appartements sont des expériences aléatoires indépendantes les unes des autres. On appelle A la variable aléatoire égale au nombre d'appartement vendus en un mois après une visite.

1. Justifier que la variable aléatoire A suit une loi binomiale dont on précisera les paramètres n et p.
2. On souhaite calculer la probabilité que l'agent vende exactement 30 appartements en un mois après une visite.
   a) Calculer C_{120}^{30} à l'aide de la calculatrice. Que peut-on conclure .
   b) Justifier que l'on peut approcher la loi de A par une loi normale que l'on précisera.
   c) À l'aide de cette approximation, calculer la probabilité de vendre exactement 30 appartements en un mois après une visite.

Réponses :

1. Il s'agit d'une expérience aléatoire admettant deux issues :
   - succès : "vendre un appartement suite à une visite" de probabilité $p = 0.15$.
   - échec : "ne pas vendre un appartement suite à une visite" de probabilité $1 - p = 1 - 0.15 = 0.85$.

   Cette expérience aléatoire est une épreuve de Bernoulli. On répète cette expérience 120 fois de manière indépendante. Donc, c'est un schéma de Bernoulli.
   On note A la variable aléatoire associé au nombre d'appartement vendus après une visite en un mois, c'est-à-dire au nombre de succès obtenus à l'issue des 120 épreuves de Bernoulli. Par conséquent A suit la loi binomiale $\mathcal{B}(n = 120; p = 0.15)$.

TD n°2 sur Moodle

- On souhaite calculer la probabilité que l'agent vende exactement 30 appartements en un mois après une visite.
  a) Calculer C_{120}^{30} à l'aide de la calculatrice. Que peut-on conclure .
  À l'aide de la calculatrice, one gets :
  $$C_{120}^{30} = 1.69745387607974 \times 10^{14}.$$
  Ce nombre est difficile à obtenir car il fait intervenir des factorielles importantes et la calculatrice nous donne qu'une valeur approchée (en réalité $C_{120}^{30} = 169745387607974089094600740 96$). De même, pour les valeurs 0.15^{30} et $(1-0.07)^{90}$ qui interviennent dans le calcul de $P(A=30)$. Pour cela, nous proposons d'approcher la loi de A par une loi convenable.
  b) Justifier que l'on peut approcher la loi de A par une loi normale que l'on précisera.
  Comme $n = 120 \geq 20$, $np = 120 \times 0.15 \geq 10$ et $n(1-p) = 120 \times 0.85 = 85 \geq 10$, on peut donc approcher la loi de A par la loi normale $\mathcal{N}(18; \sqrt{15.30}^2)$ où $E(A) = 120 \times 0.15 = 18$ et $\sigma_A = \sqrt{120 \times 0.15 \times (1-0.15)} = \sqrt{15.30}$.
  c) À l'aide de cette approximation, calculer la probabilité de vendre exactement 30 appartements en un mois après une visite.
  Soit Y une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0;1)$. On a
  $$\begin{aligned} P(A=30) = P(29.5 \leq A \leq 30.5) &= P\left(\frac{29.5-18}{\sqrt{15.30}} \leq \frac{A-18}{\sqrt{15.30}} \leq \frac{30.5-18}{\sqrt{15.30}}\right) \\ &= P\left(\frac{11.50}{\sqrt{15.30}} \leq Y \leq \frac{12.50}{\sqrt{15.30}}\right) = \Phi\left(\frac{12.50}{\sqrt{15.30}}\right) - \Phi\left(\frac{11.50}{\sqrt{15.30}}\right) \\ &\approx \Phi(3.20) - \Phi(2.94) \approx 0.99931 - 0.99836 \approx 0.00095 = 9.5 \times 10^{-4}. \end{aligned}$$

Résumé de la loi normale

Essentiel à retenir

1. La loi normale est symétrique, sa courbe est continue et sa forme est en cloche.
2. $X \hookrightarrow \mathcal{N}(0;1)$, alors $E(X) = 0$ et $V(X) = 1$.
3. $Y \hookrightarrow \mathcal{N}(\mu;\sigma^2)$, alors $X = \frac{Y-\mu}{\sigma} \hookrightarrow \mathcal{N}(0;1)$
4. $Y \hookrightarrow \mathcal{N}(\mu;\sigma^2)$, alors $E(Y) = \mu$ et $V(Y) = \sigma^2$.
5. Formules de calcul des probabilités via la table de $\mathcal{N}(0;1)$, voir la section concernée.
6. Approximation de la loi binomiale par la loi normale : soit $X \hookrightarrow \mathcal{B}(n;p)$. Si
   - $n \geq 20$;
   - $np \geq 10$.
   - $n(1-p) \geq 10$;

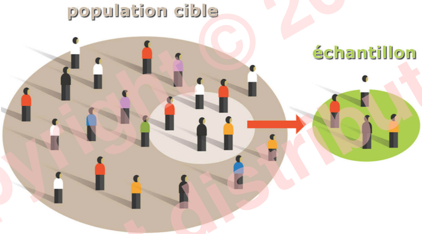
   $\rightsquigarrow$ on peut estimer la loi de X par la loi normale telle que $Y \hookrightarrow \mathcal{N}(\mu;\sigma^2)$ avec $\mu = np$, $\sigma = \sqrt{np(1-p)}$.
7. Correction de la continuité : si $n \geq 20$, $np \geq 10$ et $n(1-p) \geq 10$, alors

$$P(X = k) \approx P(k - 0.5 \leq Y \leq k + 0.5).$$

Chapitre 3 : Échantillonnage et estimation

Notions de la statistique inférentielle

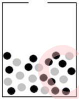
1. Notions de la statistique inférentielle
   - **Population :** ensemble des unités statistiques (personnes ou objets) aux quelles on s'intéresse et sur lesquelles porte une étude.
   - **Individu :** entité ou chaque élément de la population pour lequel les données sont collectées.
   - **Variable aléatoire parente :** caractère étudié dans la population (chiffre d'affaire, salaire, taille, ...)
   - **Échantillon :** une partie (sous ensemble) représentative de la population cible.



2. Pourquoi travaille-t-on sur un échantillon ?
   - Management : impossible de gérer toute la population, ...
   - Coût : le travail avec une population totale coûte trop cher et/ou prendre trop de temps.

Échantillonnage et estimation

- **L'échantillonnage** est une méthode statistique qui permet de passer (de la loi connue d'un paramètre dans une population cible de taille N) à une estimée d'une quantité θ_n fabriquée à partir seulement d'un échantillon de la population cible de taille $n < N$.
- **L'estimation** est une méthode statistique qui permet d'induire, à partir des résultats observées sur un échantillon, des informations sur la population cible.
- **Applications :**
  - prise de décision à partir de la proportion/fréquence par rapport à un modèle de référence.
  - estimation d'une proportion.
  - estimation de la moyenne et de la variance.
  - Définition d'une région dans laquelle un paramètre, a de grandes chances de se trouver.

Échantillonage	**Estimation**
On **connaît la proportion** p de boules rouges	On **ignore la proportion** p des boules rouges
	
On réalise des tirages avec remise de n boules et on observe la fréquence f d'apparition d'une boule rouge. f appartient à un intervalle **de fluctuation** de centre p.	On réalise des tirages avec remise de n boules pour essayer d'**estimer la proportion** p de boules rouges. On cherche un intervalle de **confiance**.

Échantillonnage

Urne U_1 : on répète indépendamment n fois la même expérience aléatoire et on appelle f la fréquence observée d'un caractère. On note p la proportion **connue** des boules rouges dans la population (urne U_1). Alors :

Propriétés

- l'intervalle de fluctuation au seuil de 90% est : $\left[p - 1.64\sqrt{\frac{p(1-p)}{n}} ; p + 1.64\sqrt{\frac{p(1-p)}{n}}\right]$.
  $\rightsquigarrow f$ appartient à cet intervalle avec un risque d'erreur de 10%.
- l'intervalle de fluctuation au seuil de 95% est : $\left[p - 1.96\sqrt{\frac{p(1-p)}{n}} ; p + 1.96\sqrt{\frac{p(1-p)}{n}}\right]$
  On dit aussi que f appartient à cet intervalle avec un risque d'erreur de 5%.
- l'intervalle de fluctuation au seuil de 99% est :
  $$\left[p - 2.58\sqrt{\frac{p(1-p)}{n}} ; p + 2.58\sqrt{\frac{p(1-p)}{n}}\right]$$
  On dit aussi que f appartient à cet intervalle avec un risque d'erreur de 1%.

Remarque : plus n est grand, plus l'amplitude de l'intervalle de fluctuation est réduite et donc on obtient un meilleur encadrement de la fréquence f.

Applications

Application n° 1 : gestion des factures

Après une enquête approfondie des factures des années précédentes, une entreprise a déterminé que le montant moyen d'une facture est de 2490€et l'écart-type de 890€. Elle tire un échantillon de 225 factures et trouve un montant moyen de 2590€.

1. Que peut-on en conclure avec un seuil de confiance de 99% ou un risque d'erreur 5% ?.
2. Quelle doit être la taille de l'échantillon pour que la moyenne soit comprise entre 2430 et 2650 avec un seuil de confiance de 99% ?

Application n° 2 : contrôle de qualité ISO

Dans le cadre d'un contrôle de qualité ISO On contrôle une production en série en tirant des échantillons de taille n. On suppose que le taux acceptable de rebut est de 15%. Sur un échantillon de 200, on trouve 35 pièces défectueuses.

1. Quelle conclusion peut-on en tirer avec un seuil de confiance de 95% ou risque d'erreur 5% ?

TD n°3 sur Moodle

Exercice 1 : On dispose d'une urne contenant un très grand nombre de boules rouges et de boules noires. On sait que la proportion de boules rouges dans l'urne est égale à 0.5. Déterminer l'intervalle de fluctuation de la fréquence d'apparition des boules rouges.

Réponses :

1. au seuil 90% pour 200 tirages successifs avec remise effectués dans l'urne.
   Il suffit d'utiliser la définition de l'intervalle de fluctuations pour la proportion $p = 0.5$ et $n = 200$. On a donc

$$I_f(90\%) = [0.5 - 1.64\sqrt{\frac{0.5(1-0.5)}{200}}; 0.5 + 1.64\sqrt{\frac{0.5(1-0.5)}{200}}] = [0.442; 0.558].$$

2. au seuil 95% pour 200 tirages successifs avec remise effectués dans l'urne.

$$I_f(95\%) = [0.5 - 1.96\sqrt{\frac{0.5(1-0.5)}{200}}; 0.5 + 1.96\sqrt{\frac{0.5(1-0.5)}{200}}] = [0.431; 0.569].$$

3. au seuil 99% pour 200 tirages successifs avec remise effectués dans l'urne.

$$I_f(99\%) = [0.5 - 2.58\sqrt{\frac{0.5(1-0.5)}{200}}; 0.5 + 2.58\sqrt{\frac{0.5(1-0.5)}{200}}] = [0.409; 0.591].$$

Estimation de la moyenne μ

Urne U_2 : on répète indépendamment n fois la même expérience aléatoire et on appelle m la moyenne observée des boules rouges. On appelle s une estimation de la population telle que

$$s = \sqrt{\frac{n}{n-1} \times (\text{moyenne des carrées des valeurs} - m^2}.$$

On note μ la moyenne **inconnue** des boules rouges dans la population (urne U_2). Alors :

Propriétés : estimation de la moyenne

- l'intervalle de confiance de μ au seuil de 90% est : $\left[m - 1.64\frac{s}{\sqrt{n}} ; m + 1.64\frac{s}{\sqrt{n}}\right]$. $\rightsquigarrow$ μ appartient à cet intervalle avec un risque d'erreur de 10%.
- l'intervalle de confiance de μ au seuil de 95% est : $\left[m - 1.96\frac{s}{\sqrt{n}} ; m + 1.96\frac{s}{\sqrt{n}}\right]$. $\rightsquigarrow$ μ appartient à cet intervalle avec un risque d'erreur de 5%.
- l'intervalle de confiance de μ au seuil de 90% est : $\left[m - 2.58\frac{s}{\sqrt{n}} ; m + 2.58\frac{s}{\sqrt{n}}\right]$. $\rightsquigarrow$ μ appartient à cet intervalle avec un risque d'erreur de 1%.

Remarque : plus n est grand, plus l'amplitude de l'intervalle de confiance est réduite et donc on obtient une meilleure approximation de la moyenne μ.

TD n°2 sur Moodle

Exercice 2 : Dans une entreprise fabriquant un grand nombre de masques chirurgicaux, on ignore la moyenne m des épaisseurs de ces masques et on n'est pas en mesure d'émettre une hypothèse quant à cette valeur. On a prélevé 200 masques successivement avec remise et on a relevé l'épaisseur de chacun d'entre eux. On a obtenu que la moyenne des épaisseurs des masques prélevés est 3.5 mm et que la moyenne des carrés des épaisseurs est égale à 12.4.
Déterminer l'intervalle de confiance de la moyenne μ :

Réponses : Pour répondre à cette question, on calcule d'abord l'estimation de l'écart-type de la population par la formule suivante :

$$s = \sqrt{\frac{n}{n-1} \times (\texttt{moyenne des carées des valeurs} - m)^2},$$

ensuite, on utilise la définition de l'intervalle de confiance au niveau donné.
Pour $n = 200$ et $m = 3.5$ et la moyenne des carrés des épaisseurs égale à 12.4, on a $s = \sqrt{\frac{200}{199} \times (12.4 - 3.5)^2} = \sqrt{\frac{30}{199}} \approx 0.388.$

1. au niveau de confiance 90%.

$$I_c(90\%) = [3.5 - 1.64 \times \frac{0.388}{\sqrt{200}}; 3.5 + 1.64 \times \frac{0.388}{\sqrt{200}}] = [3.455; 3.545].$$

2. au niveau de confiance 95%.

$$I_c(95\%) = [3.5 - 1.96 \times \frac{0.388}{\sqrt{200}}; 3.5 + 1.96 \times \frac{0.388}{\sqrt{200}}] = [3.446; 3.554].$$

Estimation de la proportion p

Urne U_2 : on répète indépendamment n fois la même expérience aléatoire et on appelle f la fréquence observée des boules rouges. On note p la proportion **inconnue** des boules rouges dans la population (urne U_2). Alors :

Propriétés : estimation de la proportion

- l'intervalle de confiance de p au seuil de 90% est : $\left[f - 1.64\sqrt{\frac{f(1-f)}{n}}; f + 1.64\sqrt{\frac{f(1-f)}{n}}\right]$. $\rightsquigarrow$ p appartient à cet intervalle avec un risque d'erreur de 10%.
- l'intervalle de confiance de p au seuil de 95% est : $\left[p - 1.96\sqrt{\frac{f(1-f)}{n}}; f + 1.96\sqrt{\frac{f(1-f)}{n}}\right]$. $\rightsquigarrow$ p appartient à cet intervalle avec un risque d'erreur de 5%.
- l'intervalle de confiance de p au seuil de 90% est : $\left[f - 2.58\sqrt{\frac{f(1-f)}{n}}; m + 2.58\sqrt{\frac{f(1-f)}{n}}\right]$. $\rightsquigarrow$ p appartient à cet intervalle avec un risque d'erreur de 1%.

Remarque : plus n est grand, plus l'amplitude de l'intervalle de confiance est réduite et donc on obtient une meilleure approximation de la proportion p.

TD n°3 sur Moodle

Exercice 3 : On dispose d'une urne contenant un très grand nombre de boules rouges et de boules noires. On ignore la proportion p de boules rouges dans l'urne et on n'est pas en mesure d'émettre une hypothèse quant à cette valeur. On réalise 500 tirages successifs avec remise dans cette urne et on obtient 389 boules noires.
Déterminer l'intervalle de confiance de p :
Puisqu'on a tiré 389 boules noires, on a tiré $500-389=$ boules rouges, si bien que $f=\frac{111}{500}=0.222$. Maintenant, on utilise la définition de l'intervalle de confiance d'une proportion p.

1. au niveau de confiance 90%.

$$I_p(90\%)=[0.222-1.64\times\sqrt{\frac{0.222\times(1-0.222)}{500}};0.222+1.64\times\sqrt{\frac{0.222\times(1-0.222)}{500}}]=[30.191;0.252]$$

2. au niveau de confiance 95%.

$$I_p(95\%)=[0.222-1.96\times\sqrt{\frac{0.222\times(1-0.222)}{500}};0.222+1.96\times\sqrt{\frac{0.222\times(1-0.222)}{500}}]=[0.185;0.258].$$

3. au niveau de confiance 99%.

$$I_p(99\%)=[0.222-2.58\times\sqrt{\frac{0.222\times(1-0.222)}{500}};0.222+2.58\times\sqrt{\frac{0.222\times(1-0.222)}{500}}]=[0.174;0.270].$$

Résumé de l'échantillonnage et l'estimation

Essentiel à retenir

- À quoi ça sert l'échantillonnage et l'estimation.
- Échantillonnage : définition d'intervalle de fluctuation au risque d'erreur $1-\alpha$
  - lorsqu'on connait la proportion p.
  - lorsqu'on émet une hypothèse sur la valeur de p (prise de décision).
- Estimation : définition d'intervalle de confiance au risque d'erreur $1-\alpha$
  - Estimation de la moyenne μ : lorsqu'on ignore la moyenne μ et lorsqu'on n'émet pas d'hypothèse sur μ.
  - Estimation de la fréquence f : lorsqu'on ignore la proportion p et lorsqu'on n'émet pas d'hypothèse sur p.