

Cours Statistiques L2
Université Nice Sophia-Antipolis

François Delarue

Table des matières

Chapitre 1. Rappels de Probabilités	5
1. Espaces de probabilité et Variables aléatoires	5
2. Espérances et variances	7
Chapitre 2. Modèles Statistiques et Estimateurs. (Rappels de Probabilités.)	11
1. Notion de Modèle Statistique	11
2. Notion d'estimateur	17
Chapitre 3. Utilisation et construction d'Estimateurs.	23
1. Intervalles et Régions de Confiance	23
2. Estimation empirique	27
Chapitre 4. Chapitre 4. Modèles Gaussiens.	37
1. Rappels : Distribution Gaussienne	37
2. Construction d'intervalles de confiance pour la moyenne	39
3. Construction d'intervalle de confiance pour l'écart-type	44
Chapitre 5. Chapitre 5. Tests.	47
1. Principes	47
2. Notions générales et autres exemples	49
3. Tests unilatères	51

Rappels de Probabilités

1. Espaces de probabilité et Variables aléatoires

1.1. Espaces de probabilité. Un espace de probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ modélisant une ou plusieurs expériences :

- (1) Ω désigne un univers contenant l'ensemble des issues possibles de la ou des expériences. (Dans le cas du lancer de dé, $\Omega = \{1, \dots, 6\}^2$.)
- (2) \mathcal{A} désigne une collection de parties de Ω , décrivant les événements observables à la suite de la ou des expériences modélisées.
- (3) \mathbb{P} est une mesure de probabilité permettant de mesurer la taille des événements.

Exemple 1. (Cas fini.) Lorsque Ω est de cardinal fini, \mathcal{A} est usuellement choisie égale à $\mathcal{P}(\Omega)$. Le cas échéant, \mathbb{P} est une mesure de probabilité s'il s'agit d'une application de $\mathcal{P}(\Omega)$ à valeurs dans $[0, 1]$ vérifiant

$$(1.1) \quad \mathbb{P}(\Omega) = 1 \quad \text{et} \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \quad A \cap B = \emptyset.$$

Exemple 2. (Cas dénombrable.) Lorsque Ω est dénombrable, \mathcal{A} est aussi usuellement choisie comme l'ensemble des parties. Par ailleurs, l'axiome (1.5) est renforcé :

$$(1.2) \quad \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n), \quad A_i \cap A_j = \emptyset \text{ si } i \neq j.$$

Exemple 3. (Cas non dénombrable.) Le problème est beaucoup plus compliqué. La collection \mathcal{A} est rarement égale à l'ensemble des parties. Elle est supposée contenir Ω , être stable par passage au complémentaire et par réunion dénombrable. Une mesure de probabilité est une application, de \mathcal{A} dans $[0, 1]$, vérifiant (1.5) et (1.4).

1.2. Variables aléatoires. Une variable aléatoire permet de décrire une expérience ou un phénomène spécifique. Précisément, il s'agit d'une application de Ω dans \mathbb{R} telle que, pour tous $a, b \in \mathbb{R} \cup \{\pm\infty\}$, les ensembles $\{\omega : X(\omega) \in (a, b)\}$ (ici, les parenthèses doivent être comprises comme $]$ ou $]$) sont dans \mathcal{A} , i.e. peuvent être mesurés.

En pratique, nous noterons $\{X \in (a, b)\}$ au lieu de $\{\omega : X(\omega) \in (a, b)\}$.

La loi d'une variable aléatoire décrit le hasard selon lequel se répartissent les issues de l'expérience considérée. Elle est donnée par l'application

$$\mathbb{P}_X : (a, b) \text{ intervalle de } \mathbb{R} \mapsto \mathbb{P}\{X \in [a, b]\}.$$

En pratique, il est fréquent d'oublier de préciser l'espace de probabilité et de se focaliser simplement sur une variable aléatoire de loi donnée. Par exemple, le lancer d'un dé peut être modélisé par une variable aléatoire X de loi

$$\mathbb{P}_X(\{1\}) = \mathbb{P}_X(\{2\}) = \dots = \mathbb{P}_X(\{6\}) = \frac{1}{6}.$$

En effet, lorsque la variable aléatoire est à valeurs discrètes (i.e. a au plus un nombre dénombrables d'images), il est suffisant de calculer les poids avec lesquels elle prend chacune des valeurs.

Exemple 1. Le lancer à pile ou face équilibré est modélisé par une variable aléatoire $X : \Omega \rightarrow \{0, 1\}$ de poids $(1/2, 1/2)$ sur $\{0, 1\}$, i.e.

$$\mathbb{P}_X(\{0\}) = \mathbb{P}_X(\{1\}) = \frac{1}{2}.$$

Lorsque la pièce est supposée déséquilibrée, les poids sont de la forme $(1 - p, p)$, avec p dans $]0, 1[$. La loi est appelée loi de Bernoulli de paramètre p . On note $X \sim \mathcal{B}(p)$.

Exemple 2. Le nombre d'apparitions d'un phénomène rare sur une très longue période est modélisé par une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ de loi de Poisson de paramètre $\lambda > 0$:

$$\forall k \in \mathbb{N}, \mathbb{P}_X(\{k\}) = \frac{\lambda^k}{k!} \exp(-\lambda).$$

On note $X \sim \mathcal{P}(\lambda)$.

Lorsque la variable aléatoire prend un nombre non dénombrable de valeurs, il ne suffit plus de calculer les poids des singletons pour la connaître. Il est en revanche possible de décrire la loi à travers la donnée d'une fonction de densité, c'est-à-dire d'une fonction continue (ou éventuellement continue par morceaux), positive et d'intégrale sur \mathbb{R} égale à 1.

Précisément,

Définition. Une variable $X : \Omega \rightarrow \mathbb{R}$ est dite de loi de densité f_X , f_X désignant une densité, si

$$\forall a \leq b, \mathbb{P}_X\{(a, b)\} = \int_a^b f_X(x) dx.$$

(Ici, a et b peuvent être infinis.)

Exemple 3. Nous appelons densité gaussienne centrée réduite la fonction

$$x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Il s'agit en effet d'une fonction (positive) d'intégrale égale à 1.

Une variable aléatoire dont la loi est donnée par cette densité est appelée loi gaussienne (ou normale) centrée réduite.

Plus généralement, pour deux paramètres $m \in \mathbb{R}$ et $\sigma > 0$, la fonction

$$x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right),$$

est une densité. (Preuve en cours ?) Une variable aléatoire dont la loi est donnée par cette densité est appelée loi gaussienne (ou normale) de paramètres m et σ^2 . On note $X \sim \mathcal{N}(m, \sigma^2)$.

En pratique, les variables de loi gaussienne décrivent les erreurs de mesure.

1.3. Fonction de répartition. Définition. Etant donnée une variable aléatoire X , la fonction de répartition de X est la fonction

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}\{X \leq t\}.$$

La fonction de répartition caractérise la loi de X : deux variables aléatoires de même fonction de répartition ont même loi.

Par extension, il est possible de parler de la fonction de répartition d'une loi, sans faire explicitement référence à la variable aléatoire sous-jacente.

Exemple 4. La loi exponentielle de paramètre $\lambda > 0$ est la loi de densité

$$p_\lambda : x \mapsto \lambda \mathbf{1}_{x \geq 0} \exp(-\lambda x).$$

Sa fonction de répartition est donnée par

$$F : t \mapsto \begin{cases} 1 & \text{si } t \leq 0, \\ 1 - \exp(-\lambda t) & \text{si } t > 0. \end{cases}$$

2. Espérances et variances

2.1. Espérance d'une variable de Bernoulli. Commençons par rappeler la définition de l'espérance dans le cas d'une variable indicatrice :

Définition. Etant donné un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et un événement A , on appelle indicatrice de A la variable aléatoire

$$\mathbf{1}_A : \omega \in \Omega \mapsto \begin{cases} 0 & \text{si } \omega \notin A, \\ 1 & \text{si } \omega \in A. \end{cases}$$

Il s'agit d'une variable aléatoire à valeurs dans $\{0, 1\}$ de loi de Bernoulli de paramètre $\mathbb{P}(A)$.

Par définition, son espérance, notée $\mathbb{E}[\mathbf{1}_A]$ est égale à $\mathbb{P}(A)$.

Cette définition s'interprète facilement au regard de la loi des grands nombres, rappelée plus tard. Expérimentalement, il s'agit de lancer une série de pièces à pile ou face selon un paramètre de succès p . Intuitivement, il est légitime d'espérer que le nombre moyen de succès en temps long soit proche de p : ce nombre moyen de succès s'écrit, en temps n , comme la moyenne empirique

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{A_i},$$

où A_i désigne l'événement : le i ème lancer est un succès. Cette moyenne empirique apparaît comme un gain moyenné, celui obtenu en ramassant 1 euro à chaque succès. Ici, toutes les variables $\mathbf{1}_{A_i}$, $i \geq 1$, ont même loi : leur gain théorique ou mathématique, ou encore leur espérance mathématique, est posé égal à la limite du gain moyenné au cours du temps, i.e. égal à p . Cette définition, remarquable, consiste à identifier une moyenne au cours du temps avec une moyenne sur un espace. En effet, $\mathbb{E}[\mathbf{1}_A]$ peut se comprendre comme

$$\mathbb{E}[\mathbf{1}_A] = 0 \times \mathbb{P}\{\mathbf{1}_A = 0\} + 1 \times \mathbb{P}\{\mathbf{1}_A = 1\},$$

i.e. comme un barycentre.

2.2. Cas des variables discrètes. Dans le cas des variables aléatoires discrètes (i.e. avec un nombre au plus dénombrable de valeurs), la définition barycentrique demeure :

Définition. Etant donnée une variable aléatoire X à valeurs dans un ensemble fini $\{x_1, \dots, x_n\}$, l'espérance de X est donnée :

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}\{X = x_i\}.$$

Lorsque X est à valeurs dans un ensemble dénombrable $\{(x_n), n \geq 1\}$, il est nécessaire de vérifier que la série est convergente, i.e. que le barycentre est bien défini. Pour des raisons un peu techniques, l'espérance n'est considérée comme définie que lorsque la série est absolument convergente (rappelons par exemple que la valeur d'une série semi-convergente est sensible à l'ordre de la sommation, de sorte que la notion de barycentre n'a, le cas échéant, aucun sens canonique). De fait, si

$$\sum_{n \geq 1} |x_n| \mathbb{P}\{X = x_n\} < +\infty,$$

nous posons

$$\mathbb{E}(X) = \sum_{n \geq 1} x_n \mathbb{P}\{X = x_n\}.$$

Remarquons que l'espérance ne dépend que de la loi de la variable aléatoire. (De fait, nous parlerons par extension de l'espérance d'une loi.)

Exemple 1. Rappelons que la loi binomiale de paramètres n et p est la loi sur $\{0, \dots, n\}$ de poids

$$p_k = C_n^k p^k (1-p)^{n-k}.$$

Alors, l'espérance de la loi binomiale est donnée par np . (Preuve en cours?)

Exemple 2. L'espérance de la loi de Poisson de paramètre λ est λ .

Cette notion barycentrique se généralise au cas de la composée d'une variable X et d'une fonction f de \mathbb{R} dans \mathbb{R} .

Proposition. Soient X une variable à valeurs dans un ensemble dénombrable $\{(x_n)_{n \geq 1}\}$ et f une fonction de \mathbb{R} dans \mathbb{R} telle que

$$\sum_{n \geq 1} |f(x_n)| \mathbb{P}\{X = x_n\} < +\infty,$$

alors,

$$\mathbb{E}[f(X)] = \sum_{n \geq 1} f(x_n) \mathbb{P}\{X = x_n\}.$$

(La définition se généralise de façon triviale au cas fini.)

2.3. Espérance d'une variable de loi à densité. La définition généralise la notion de barycentre à travers celle d'intégrale.

Définition-Proposition. Soit X une variable de loi de densité g (continue ou continue par morceaux) telle que

$$\int_{\mathbb{R}} |x|g(x)dx < +\infty.$$

Alors, l'espérance de X est posée égale à

$$\mathbb{E}(X) = \int_{\mathbb{R}} xg(x)dx.$$

Plus généralement, si f désigne une fonction de \mathbb{R} dans \mathbb{R} (continue) telle que

$$\int_{\mathbb{R}} |f(x)|g(x)dx < +\infty.$$

Alors,

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x)g(x)dx.$$

2.4. Variance. La variance permet de calculer le carré de la distance à la moyenne :

Définition-Proposition. Soit X une variable telle que $\mathbb{E}(X^2)$ soit bien définie, alors $\mathbb{E}(X)$ est bien définie et la variance de X est

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2],$$

terme également bien défini.

Modèles Statistiques et Estimateurs. (Rappels de Probabilités.)

1. Notion de Modèle Statistique

1.1. Exemples de problématique.

Problème 1. Un fabricant de lampes électriques souhaite tester la qualité d'un nouveau produit, que ses ingénieurs viennent de mettre au point, et vise plus particulièrement à en étudier la durée de vie. Pour cela, il sélectionne, au hasard, 100 de ces lampes électriques, et les branche en "parallèle" sur un circuit électrique. Il obtient un 100-uplet (t_1, \dots, t_{100}) résumant les durées de vie de chacune des ampoules, de la première (par exemple, celle montée le plus à gauche du circuit) à la dernière (dans notre exemple, celle montée le plus à droite).

Problème 2. Un fabricant de CD vierges décide de mener une étude de qualité de ses produits, visant à déterminer la proportion de CD défectueux à la sortie de la chaîne de fabrication. Il propose pour cela de tester, au hasard, un millier de CD issus de son usine et de consigner les résultats sous la forme d'un 1000-uplet $(\varepsilon_1, \dots, \varepsilon_{1000})$ à valeurs dans $\{0, 1\}$, 0 signifiant la défectuosité du CD testé et 1 son aptitude à la vente.

Ces deux problèmes, en apparence différents, soulèvent en réalité des questions et des enjeux similaires :

- (1) Dans les deux cas, l'objectif consiste, au moins pour partie, à estimer une quantité d'intérêt : durée de vie de l'ampoule dans le premier problème et proportion de produits défectueux dans le deuxième.
- (2) Dans les deux cas, l'estimation est envisagée à partir de l'observation d'une partie de la population et non de sa totalité, ou dit autrement, d'un échantillon de la population.
- (3) Dans les deux cas, l'observation de la population est non seulement susceptible de permettre d'estimer les quantités d'intérêt mais aussi de participer à une aide à la décision : il peut par exemple s'agir de décider si l'ampoule mise au point répond ou non à un cahier des

charges préalablement établi ou si la fiabilité de la chaîne de production est à la hauteur du plan de développement de l'usine. Dans ces deux cas, l'aide à la décision s'apparente à une forme de test d'une hypothèse initiale, validant un acquis ("l'ampoule est conforme au cahier des charges" ou "la chaîne de production est suffisamment fiable"), contre une hypothèse alternative, aux conséquences éventuellement lourdes (remise en cause du produit ou de la chaîne de fabrication).

La théorie des statistiques vise à donner un cadre rigoureux permettant à la fois de formaliser les expériences pratiques, de préciser la qualité des estimations effectuées à l'issue des expériences et enfin de suggérer des tests favorisant la prise de décision.

1.2. Modélisation de l'expérience. Dans les deux problèmes proposés, l'expérience repose sur l'observation de phénomènes aléatoires comme la durée de vie d'une ampoule ou la qualité d'un CD à l'issue de la chaîne de production. Il s'agit de fait de proposer un cadre rigoureux permettant de décrire l'expérience. (En l'espèce, l'expérience ici peut être comprise comme une famille d'expériences : il s'agit par exemple de mesurer la durée de chacune des ampoules.)

En probabilité, la modélisation d'une famille d'expériences s'appuie sur la notion d'espace de probabilité. Un espace de probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ modélisant une ou plusieurs expériences :

- (1) Ω désigne un univers contenant l'ensemble des issues possibles de la ou des expériences. (Dans le cas du lancer de dé, $\Omega = \{1, \dots, 6\}^2$.)
- (2) \mathcal{A} désigne une collection de parties de Ω , décrivant les événements observables à la suite de la ou des expériences modélisées.
- (3) \mathbb{P} est une mesure de probabilité permettant de mesurer la taille des événements.

Exemple 1. (Cas fini.) Lorsque Ω est de cardinal fini, \mathcal{A} est usuellement choisie égale à $\mathcal{P}(\Omega)$. Le cas échéant, \mathbb{P} est une mesure de probabilité s'il s'agit d'une application de $\mathcal{P}(\Omega)$ à valeurs dans $[0, 1]$ vérifiant

$$(1.3) \quad \mathbb{P}(\Omega) = 1 \quad \text{et} \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \quad A \cap B = \emptyset.$$

Exemple 2. (Cas dénombrable.) Lorsque Ω est dénombrable, \mathcal{A} est aussi usuellement choisie comme l'ensemble des parties. Par ailleurs, l'axiome (1.5) est renforcé :

$$(1.4) \quad \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n), \quad A_i \cap A_j = \emptyset \text{ si } i \neq j.$$

Exemple 3. (Cas non dénombrable.) Le problème est beaucoup plus compliqué. La collection \mathcal{A} est rarement égale à l'ensemble des parties. Elle est supposée contenir Ω , être stable par passage au complémentaire et par réunion dénombrable. Une mesure de probabilité est une application, de \mathcal{A} dans $[0, 1]$, vérifiant (1.5) et (1.4).

1.3. Description d'une observation (ou d'une expérience spécifique). Répétons-le si nécessaire : l'espace de probabilité donne un cadre rigoureux à la modélisation d'une famille d'expériences. En pratique, il peut être nécessaire de décrire, par un objet mathématique précis, une expérience donnée ou une observation spécifique réalisée au cours du processus expérimental : par exemple, il peut s'agir de décrire la durée de vie de la 55ème ampoule testée ou la qualité du 434ème produit tiré au hasard.

En probabilités, la description d'un trait ou d'une expérience spécifique s'appuie sur la notion de variable aléatoire. Précisément, une variable aléatoire est une application de Ω dans \mathbb{R} telle que, pour tous $a, b \in \mathbb{R} \cup \{\pm\infty\}$, les ensembles $\{\omega : X(\omega) \in (a, b)\}$ (ici, les parenthèses doivent être comprises comme $]$ ou $[)$ sont dans \mathcal{A} , i.e. peuvent être mesurés.

En pratique, nous noterons $\{X \in (a, b)\}$ au lieu de $\{\omega : X(\omega) \in (a, b)\}$.

La loi d'une variable aléatoire décrit le hasard selon lequel se répartissent les issues de l'expérience considérée. Elle est donnée par l'application

$$\mathbb{P}_X : (a, b) \text{ intervalle de } \mathbb{R} \mapsto \mathbb{P}\{X \in [a, b]\}.$$

En pratique, il est également fréquent d'oublier de préciser l'espace de probabilité et de se focaliser simplement sur une variable aléatoire de loi donnée. Par exemple, le lancer d'un dé peut être modélisé par une variable aléatoire X de loi

$$\mathbb{P}_X(\{1\}) = \mathbb{P}_X(\{2\}) = \dots = \mathbb{P}_X(\{6\}) = \frac{1}{6}.$$

Au passage, cet exemple rappelle que, lorsque la variable aléatoire est à valeurs discrètes (i.e. a au plus un nombre dénombrables d'images), il est suffisant de calculer les poids avec lesquels elle prend chacune des valeurs.

Exemple 4. Le lancer à pile ou face équilibré est modélisé par une variable aléatoire $X : \Omega \rightarrow \{0, 1\}$ de poids $(1/2, 1/2)$ sur $\{0, 1\}$, i.e.

$$\mathbb{P}_X(\{0\}) = \mathbb{P}_X(\{1\}) = \frac{1}{2}.$$

Lorsque la pièce est supposée déséquilibrée, les poids sont de la forme $(1 - p, p)$, avec p dans $]0, 1[$. La loi est appelée loi de Bernoulli de paramètre p . On note $X \sim \mathcal{B}(p)$.

Exemple 5. Le nombre d'apparitions d'un phénomène rare sur une très longue période est modélisé par une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ de loi de

Poisson de paramètre $\lambda > 0$:

$$\forall k \in \mathbb{N}, \mathbb{P}_X(\{k\}) = \frac{\lambda^k}{k!} \exp(-\lambda).$$

On note $X \sim \mathcal{P}(\lambda)$.

Lorsque la variable aléatoire prend un nombre non dénombrable de valeurs, il ne suffit plus de calculer les poids des singletons pour la connaître. Il est en revanche possible de décrire la loi à travers la donnée d'une fonction de densité, c'est-à-dire d'une fonction continue (ou éventuellement continue par morceaux), positive et d'intégrale sur \mathbb{R} égale à 1.

Précisément,

Définition 1. Une variable $X : \Omega \rightarrow \mathbb{R}$ est dite de loi de densité f_X , f_X désignant une densité, si

$$\forall a \leq b, \mathbb{P}_X\{(a, b)\} = \int_a^b f_X(x) dx.$$

(Ici, a et b peuvent être infinis.)

Exemple 6. Nous appelons densité gaussienne centrée réduite la fonction

$$x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Il s'agit en effet d'une fonction (positive) d'intégrale égale à 1.

Une variable aléatoire dont la loi est donnée par cette densité est appelée loi gaussienne (ou normale) centrée réduite.

Plus généralement, pour deux paramètres $m \in \mathbb{R}$ et $\sigma > 0$, la fonction

$$x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right),$$

est une densité. (Preuve en cours ?) Une variable aléatoire dont la loi est donnée par cette densité est appelée loi gaussienne (ou normale) de paramètres m et σ^2 . On note $X \sim \mathcal{N}(m, \sigma^2)$.

En pratique, les variables de loi gaussienne décrivent les erreurs de mesure.

Rappelons enfin que la loi d'une variable aléatoire peut être résumée par :

Définition 2. Etant donnée une variable aléatoire X , la fonction de répartition de X est la fonction

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}\{X \leq t\}.$$

La fonction de répartition caractérise la loi de X : deux variables aléatoires de même fonction de répartition ont même loi. Par extension, il est possible

de parler de la fonction de répartition d'une loi, sans faire explicitement référence à la variable aléatoire sous-jacente.

Exemple 7. La loi exponentielle de paramètre $\lambda > 0$ est la loi de densité

$$p_\lambda : x \mapsto \lambda \mathbf{1}_{x>0} \exp(-\lambda x).$$

Sa fonction de répartition est donnée par

$$F : t \mapsto \begin{cases} 0 & \text{si } t \leq 0, \\ 1 - \exp(-\lambda t) & \text{si } t > 0. \end{cases}$$

1.4. Modèle statistique. Nous avons maintenant (presque) tous les outils pour modéliser les procédures expérimentales 1 et 2 introduites en début de chapitre.

Dans le premier problème, nous comprenons que les mesures des durées de vie des 100 ampoules testées sont à représenter par les valeurs de 100 variables aléatoires T_1, \dots, T_{100} . Autrement dit, si le 100-uplet (t_1, \dots, t_{100}) désigne les valeurs expérimentales consignées à l'issue du processus de mesure, nous écrivons (t_1, \dots, t_{100}) comme la réalisation de (T_1, \dots, T_{100}) pour une issue ω du hasard (ou encore pour une issue de la famille d'expériences), i.e.

$$(t_1, \dots, t_{100}) = (T_1(\omega), \dots, T_{100}(\omega)).$$

Les ampoules testées résultant du même schéma de fabrication, il est légitime de supposer que le hasard régissant leurs durées de vie se distribue en réalité de la même façon pour chacune. Autrement dit, nous supposons que les variables aléatoires T_1, \dots, T_{100} ont même loi. Enfin, avant même d'avoir rappelé la notion d'indépendance de variables aléatoires, nous comprenons que, les ampoules ayant été choisies au hasard, leurs durées de vie sont indépendantes.

Un raisonnement similaire permettrait de modéliser les observations $(\varepsilon_1, \dots, \varepsilon_{1000})$ du contrôle qualité à l'aide de 1000 variables $(\mathcal{E}_1, \dots, \mathcal{E}_{1000})$ indépendantes et, plus spécifiquement, d'écrire $(\varepsilon_1, \dots, \varepsilon_{1000})$ comme $(\mathcal{E}_1(\omega), \dots, \mathcal{E}_{1000}(\omega))$, pour une issue $\omega \in \Omega$.

Avons-nous tout dit ? Non, il reste le plus important : choisir la probabilité \mathbb{P} . Ou, de façon différente : choisir la loi de T_1, \dots, T_{100} ou de $\mathcal{E}_1, \dots, \mathcal{E}_{1000}$.

Par exemple, dans le cas du contrôle qualité, nous comprenons que chacune des variables aléatoires suit une loi de Bernoulli. En revanche, nous sommes incapables de dire quelle est la valeur du paramètre : le paramètre de la Bernoulli décrit exactement la proportion d'objets non-défectueux dans la population, que l'expérience vise justement à estimer. Nous sommes de fait réduits à choisir \mathbb{P} parmi une famille de mesures de probabilité, indexée par la valeur inconnue du paramètre $p \in [0, 1]$.

Concernant la durée de vie des ampoules, le problème est un peu plus compliqué : à première vue, aucune famille de lois ne s'impose aussi facilement que dans le cas du contrôle qualité. En fait, il est de coutume de modéliser la durée de vie d'une ampoule (ou plus généralement d'un composant électronique) par une loi exponentielle. Mais, là encore, nous sommes incapables de dire quelle est la valeur du paramètre de la loi exponentielle régissant le modèle : la connaissance du paramètre implique la connaissance de la durée de vie moyenne, que l'expérience vise justement à estimer. A nouveau, nous sommes de fait réduits à choisir \mathbb{P} parmi une famille de mesures de probabilité, indexée par la valeur inconnue du paramètre de la loi exponentielle, à valeurs dans $]0, +\infty[$.

Définition 3. Un modèle statistique paramétré est une famille d'espaces de probabilité $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ où θ est un paramètre évoluant dans un intervalle Θ (ou plus généralement dans un ouvert ou un fermé de \mathbb{R}^d , $d \geq 1$).

Ici, Ω décrit les issues possibles d'une "grosse" expérience, \mathcal{A} les événements observables à l'issue de cette "grosse" expérience et \mathbb{P}_θ une mesure susceptible de décrire la réalité, mais susceptible seulement. (En réalité, une mesure au moins est supposée décrire la réalité.)

Définition 4. Un modèle statistique échantillonné de taille n et de lois $(\mu_\theta)_{\theta \in \Theta}$, où μ_θ est une loi de probabilité (discrète ou à densité) pour chaque θ , est un modèle statistique de la forme $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$, muni de n applications X_1, \dots, X_n , telles que :

- (1) X_1, \dots, X_n sont des variables aléatoires,
- (2) pour chaque θ , X_1, \dots, X_n sont indépendantes sous \mathbb{P}_θ
- (3) pour chaque θ et pour chaque $i \in \{1, \dots, n\}$, la loi de X_i sous \mathbb{P}_θ est μ_θ .

Exemple 8. Dans le cas de la fabrication d'ampoules, n vaut 100, Θ peut être choisi comme $]0, +\infty[$ et μ_θ comme la loi exponentielle de paramètre θ .

Exemple 9. Dans le cas du contrôle qualité, n vaut 1000, Θ peut être choisi comme $\{0, 1\}$ et μ_θ comme la loi de Bernoulli de paramètre θ .

Notations. Un modèle statistique échantillonné de taille n et de lois $(\mu_\theta)_{\theta \in \Theta}$ est noté, de façon générique, sous la forme $(\mathbb{R}^n, \mathbb{P}_\theta = \mu_\theta^{\otimes n})_{\theta \in \Theta}$. Répétons-le : cette notation signifie que sont effectuées n observations, dans des conditions indépendantes, d'un phénomène dont la loi est un élément de la famille $(\mu_\theta)_{\theta \in \Theta}$.

Proposition 1. Etant donnée une famille de lois $(\mu_\theta)_{\theta \in \Theta}$, nous admettons l'existence d'un modèle statistique échantillonné de taille n .

Définition 5. Un modèle statistique échantillonné est dit *identifiable* si l'application $\theta \in \Theta \mapsto \mu_\theta$ est injective, i.e.

$$\mu_\theta = \mu_{\theta'} \Rightarrow \theta = \theta'.$$

(Faire en cours l'exemple de la Bernoulli.)

2. Notion d'estimateur

2.1. Statistique. En statistique, la notion de variable aléatoire est en réalité remplacée par celle de statistique : il s'agit d'un changement de vocabulaire permettant d'insister sur la multiplicité des mesures de probabilité sous-jacentes.

Définition 6. Etant donné un modèle statistique de la forme $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$, une statistique est une application de Ω dans \mathbb{R} telle que, pour tout $a \leq b$, l'image réciproque $\{X \in (a, b)\}$ (avec la même convention que précédemment sur l'ouverture et la fermeture des bornes) soit dans \mathcal{A} .

Insistons : il n'y a pas de différence, à proprement parler, avec une variable aléatoire. Le vocabulaire indique simplement la présence d'une structure statistique à la différence d'un modèle probabiliste sur lequel est fixée la mesure de probabilité.

2.2. Estimateurs. Ici, nous considérons un modèle statistique paramétré par un paramètre θ évoluant dans un intervalle Θ de \mathbb{R} ou, plus généralement, dans un ouvert ou dans un fermé de \mathbb{R}^d , $d \geq 1$. En pratique, il s'agit d'estimer le paramètre θ ou une fonction de θ , i.e. une quantité de la forme $g(\theta)$, avec $g : \Theta \rightarrow \mathbb{R}^n$.

Par exemple, dans le modèle échantillonné $(\mathbb{R}^{100}, \text{Exp}^{\otimes 100}(\lambda))$ permettant de modéliser l'estimation de la durée de vie des ampoules testées, la quantité d'intérêt est la fonction $g(\lambda) = 1/\lambda$, où $\lambda > 0$. En effet, nous savons que l'espérance d'une loi exponentielle est $1/\lambda$. (Ce point sera rappelé très bientôt.)

Nous introduisons donc la définition :

Définition 7. Etant donné un modèle statistique paramétré $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ est une fonction $g : \Theta \rightarrow \mathbb{R}^n$ (continue ou continue par morceaux), un estimateur de $g(\theta)$ est une statistique à valeurs dans \mathbb{R}^n .

Attention : un estimateur ne doit pas dépendre de θ !

Exemple 10. Le problème de qualité introduit en début de chapitre est modélisé par un modèle échantillonné de la forme $(\mathbb{R}^{1000}, \text{Ber}^{\otimes 1000}(p))_{p \in [0,1]}$. Il s'agit d'une notation pour indiquer la donnée d'un modèle statistique $(\Omega, \mathcal{A}, \mathbb{P}_p)_{p \in [0,1]}$ muni de 1000 statistiques $(\mathcal{E}_i)_{1 \leq i \leq 1000}$ à valeurs dans $\{0, 1\}$,

telles que, sous chaque \mathbb{P}_p , les variables $\mathcal{E}_1, \dots, \mathcal{E}_{1000}$ soient indépendantes et de même loi de Bernoulli de paramètre p . Un estimateur du paramètre p est une statistique (indépendante de p) à valeurs dans \mathbb{R} .

Naturellement, l'estimateur est recherché comme une fonction des statistiques d'observation $\mathcal{E}_1, \dots, \mathcal{E}_{1000}$. Par exemple, cela peut être \mathcal{E}_1 , $\mathcal{E}_1 + \mathcal{E}_2$, $\mathcal{E}_1 \times \mathcal{E}_2$. Mais, cela ne peut être ni p , ni $\mathcal{E}_1 + p \dots$

Exemple 11. Le problème de durée de vie est quant-à lui modélisé par un modèle échantillonné de la forme $(\mathbb{R}^{100}, \text{Exp}^{\otimes 100}(\lambda))_{\lambda > 0}$. Cela signifie que nous nous sommes donné un modèle statistique $(\Omega, \mathcal{A}, \mathbb{P}_\lambda)_{\lambda > 0}$ muni de 100 statistiques $(T_i)_{1 \leq i \leq 100}$ à valeurs dans \mathbb{R}_+^* , telles que, sous chaque \mathbb{P}_λ , les variables T_1, \dots, T_{100} soient indépendantes et de même loi exponentielle de paramètre λ . Un estimateur du paramètre λ est une statistique (indépendante de λ) à valeurs dans \mathbb{R} .

Là encore, l'estimateur est recherché comme une fonction des statistiques d'observation T_1, \dots, T_{100} . Par exemple, cela peut être T_1 , $1/(T_1 + \dots + T_{100})$. Mais, cela ne peut être ni λ , ni $\max(\lambda, T_{100}) \dots$

2.3. Intégrabilité des estimateurs : exemple de critères de choix.

La définition donnée en introduction est, d'apparence, un peu trop large : aucun critère, outre celui de la dimension, n'est imposé à l'estimateur... Un estimateur de $g(\theta)$ peut, de fait, être très différent de $g(\theta)$.

Une première façon de restreindre la classe d'estimateurs consiste à se focaliser sur ceux, qui en moyenne, se comportent comme $g(\theta)$. Pour cela, il est nécessaire de pouvoir effectivement calculer des moyennes :

Définition 8. Etant donné un modèle statistique $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$, une statistique S est dite intégrable si

$$\forall \theta \in \Theta, \mathbb{E}_\theta[|S|] < +\infty.$$

En particulier, étant donné une quantité d'intérêt $g(\theta)$ à estimer (g à valeurs dans \mathbb{R}) et un estimateur X , X est dit *sans biais* s'il est intégrable et

$$\forall \theta \in \Theta, \mathbb{E}_\theta[X] = g(\theta).$$

Afin de donner des exemples, nous rappelons quelques éléments de la théorie des probabilités sur la notion d'espérance.

2.4. Espérance d'une variable de Bernoulli. Commençons par rappeler la définition de l'espérance dans le cas d'une variable indicatrice :

Définition 9. Etant donné un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et un événement A , on appelle indicatrice de A la variable aléatoire

$$\mathbf{1}_A : \omega \in \Omega \mapsto \begin{cases} 0 & \text{si } \omega \notin A, \\ 1 & \text{si } \omega \in A. \end{cases}$$

Il s'agit d'une variable aléatoire à valeurs dans $\{0, 1\}$ de loi de Bernoulli de paramètre $\mathbb{P}(A)$.

Par définition, son espérance, notée $\mathbb{E}[\mathbf{1}_A]$ est égale à $\mathbb{P}(A)$.

Cette définition s'interprète facilement au regard de la loi des grands nombres, rappelée plus tard. Expérimentalement, il s'agit de lancer une série de pièces à pile ou face selon un paramètre de succès p . Intuitivement, il est légitime d'espérer que le nombre moyen de succès en temps long soit proche de p : ce nombre moyen de succès s'écrit, en temps n , comme la moyenne empirique

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{A_i},$$

où A_i désigne l'événement : le i ème lancer est un succès. Cette moyenne empirique apparaît comme un gain moyenné, celui obtenu en ramassant 1 euro à chaque succès. Ici, toutes les variables $\mathbf{1}_{A_i}$, $i \geq 1$, ont même loi : leur gain théorique ou mathématique, ou encore leur espérance mathématique, est posé égal à la limite du gain moyenné au cours du temps, i.e. égal à p . Cette définition, remarquable, consiste à identifier une moyenne au cours du temps avec une moyenne sur un espace. En effet, $\mathbb{E}[\mathbf{1}_A]$ peut se comprendre comme

$$\mathbb{E}[\mathbf{1}_A] = 0 \times \mathbb{P}\{\mathbf{1}_A = 0\} + 1 \times \mathbb{P}\{\mathbf{1}_A = 1\},$$

i.e. comme un barycentre.

2.5. Cas des variables discrètes. Dans le cas des variables aléatoires discrètes (i.e. avec un nombre au plus dénombrable de valeurs), la définition barycentrique demeure :

Définition 10. Etant donnée une variable aléatoire X à valeurs dans un ensemble fini $\{x_1, \dots, x_n\}$, l'espérance de X est donnée :

$$\mathbb{E}(X) = \sum_{i=1}^n x_i \mathbb{P}\{X = x_i\}.$$

Lorsque X est à valeurs dans un ensemble dénombrable $\{(x_n), n \geq 1\}$, il est nécessaire de vérifier que la série est convergente, i.e. que le barycentre est bien défini. Pour des raisons un peu techniques, l'espérance n'est considérée comme définie que lorsque la série est absolument convergente (rappelons par exemple que la valeur d'une série semi-convergente est sensible à l'ordre de la sommation, de sorte que la notion de barycentre n'a, le cas échéant, aucun sens canonique). De fait, si

$$\sum_{n \geq 1} |x_n| \mathbb{P}\{X = x_n\} < +\infty,$$

nous posons

$$\mathbb{E}(X) = \sum_{n \geq 1} x_n \mathbb{P}\{X = x_n\}.$$

Remarquons que l'espérance ne dépend que de la loi de la variable aléatoire. (De fait, nous parlerons par extension de l'espérance d'une loi.)

Exemple 12. Rappelons que la loi binomiale de paramètres n et p est la loi sur $\{0, \dots, n\}$ de poids

$$p_k = C_n^k p^k (1-p)^{n-k}.$$

Alors, l'espérance de la loi binomiale est donnée par np . (Preuve en cours?)

Exemple 13. L'espérance de la loi de Poisson de paramètre λ est λ .

Cette notion barycentrique se généralise au cas de la composée d'une variable X et d'une fonction f de \mathbb{R} dans \mathbb{R} .

Proposition 2. Soient X une variable à valeurs dans un ensemble dénombrable $\{(x_n)_{n \geq 1}\}$ et f une fonction de \mathbb{R} dans \mathbb{R} telle que

$$\sum_{n \geq 1} |f(x_n)| \mathbb{P}\{X = x_n\} < +\infty,$$

alors,

$$\mathbb{E}[f(X)] = \sum_{n \geq 1} f(x_n) \mathbb{P}\{X = x_n\}.$$

(La définition se généralise de façon triviale au cas fini.)

2.6. Espérance d'une variable de loi à densité. La définition généralise la notion de barycentre à travers celle d'intégrale.

Définition-Proposition 1. Soit X une variable de loi de densité g (continue ou continue par morceaux) telle que

$$\int_{\mathbb{R}} |x|g(x)dx < +\infty.$$

Alors, l'espérance de X est posée égale à

$$\mathbb{E}(X) = \int_{\mathbb{R}} xg(x)dx.$$

Plus généralement, si f désigne une fonction de \mathbb{R} dans \mathbb{R} (continue) telle que

$$\int_{\mathbb{R}} |f(x)|g(x)dx < +\infty.$$

Alors,

$$\mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x)g(x)dx.$$

2.7. Variance : vers la notion de risque. La variance permet de calculer le carré de la distance à la moyenne :

Définition-Proposition 2. Soit X une variable telle que $\mathbb{E}(X^2)$ soit bien définie, alors $\mathbb{E}(X)$ est bien définie et la variance de X est

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2],$$

terme également bien défini.

L'intérêt statistique de cette notion est aisément compréhensible. Pour un estimateur sans biais, la variance s'apparente à un risque et permet de mesurer l'écart entre l'estimateur et la quantité d'intérêt.

Définition 11. Etant donné un modèle statistique $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ et une fonction g du paramètre θ , le risque quadratique d'un estimateur sans biais T de $g(\theta)$ est donné par

$$R(\theta) = \mathbb{V}_\theta(T) = \mathbb{E}_\theta[|T - \mathbb{E}_\theta(T)|^2] = \mathbb{E}_\theta[|T - g(\theta)|^2].$$

En particulier, l'estimateur sans biais T de $g(\theta)$ est dit de moindre risque quadratique que un autre estimateur sans biais T' si

$$\forall \theta \in \Theta, \mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T').$$

CHAPITRE 3

Utilisation et construction d'Estimateurs.

Dans le chapitre précédent, nous avons introduit la notion d'estimateurs ainsi que deux règles de choix, l'une fondée sur le biais, l'autre sur le risque. Dans ce chapitre, nous cherchons :

- (1) A illustrer l'utilisation des estimateurs, particulièrement ceux répondant aux deux critères de choix que nous venons de rappeler,
- (2) à donner quelques principaux généraux de construction d'estimateurs pertinents.

1. Intervalles et Régions de Confiance

1.1. Position du problème. Dans le Chapitre 1, nous avons appuyé l'introduction des modèles statistiques sur deux problèmes pratiques :

- (1) Un modèle d'étude de la durée de vie d'une ampoule électrique, fondé sur l'observation de 100 expériences et modélisé par le modèle échantillonné

$$(\Omega, \text{Exp}^{\otimes 100}(\lambda))_{\lambda > 0}.$$

- (2) Un modèle d'étude du contrôle de la qualité d'une chaîne de fabrication, fondé sur l'observation de 1000 produits fabriqués et modélisé par le modèle échantillonné

$$(\Omega, \text{Ber}^{\otimes 1000}(p))_{p \in]0,1[}.$$

Plusieurs estimateurs ont été proposés pour chacun de ces deux modèles. Par exemple, il a été montré que la statistique canonique X_i (modélisant l'observation numéro i) est un, dans le premier cas, un estimateur sans biais de $1/\lambda$ et, dans le deuxième cas, un est un estimateur sans biais de p .

Nous pouvons également calculer le risque quadratique dans chacun des deux cas.

Dans le premier cas, nous écrivons, pour $\lambda > 0$,

$$\begin{aligned}\mathbb{E}_\lambda[(X_i - \lambda^{-1})^2] &= \mathbb{V}_\lambda(X_i) \\ &= \lambda \int_0^{+\infty} x^2 \exp(-\lambda x) dx - \lambda^{-2} \\ &= [-x^2 \exp(-\lambda x)]_0^{+\infty} + 2 \int_0^{+\infty} x \exp(-\lambda x) dx - \lambda^{-2} \\ &= \lambda^{-2}.\end{aligned}$$

Dans le deuxième cas, nous écrivons, sur le même modèle :

$$\mathbb{E}_p[(X_i - p)^2] = p(1 - p).$$

Chacune de ces quantités donne une information sur le risque liée à l'estimation de la quantité d'intérêt à l'aide d'une (et d'une seule) observation.

Cette information peut néanmoins paraître quelque peu abstraite. Nous allons essayer d'en déduire une estimation « parlante » du paramètre à estimer.

1.2. Utilisation de l'inégalité de Bienaymé-Chebychev. Nous commençons par rappeler

Lemme. (Inégalité de Bienaymé-Chebychev.) Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire de carré intégrable. Alors, en notant $m = \mathbb{E}(X)$ et $\sigma^2 = \mathbb{V}(X)$,

$$\forall a > 0, \quad \mathbb{P}\{|X - m| \geq a\} \leq \frac{\sigma^2}{a^2}.$$

Commentons : l'inégalité de Bienaymé-Chebychev relie les déviations de la variable aléatoire à sa moyenne à l'aide de la variance. Naturellement, plus l'estimation de la probabilité d'observer une déviation donnée est d'autant plus petite que la déviation considérée est grande. Soulignons également combien il est naturel de s'appuyer sur la variance pour estimer les probabilités de déviation à la moyenne : la variance quantifie justement l'écart entre les valeurs d'une variable aléatoire et sa moyenne.

Preuve. Nous appelons A l'événement $A = \{|X - m| > a\}$. Alors, nous remarquons que la variable aléatoire $\mathbf{1}_A$ est toujours plus petite que $a^{-2}|X - m|^2$. En effet, sur A , $a^{-2}|X - m|^2 \geq 1$. L'espérance de $\mathbf{1}_A$ est donc plus petite que celle de $a^{-2}|X - m|^2$. \square

Que nous apprend l'estimation de Bienaymé-Chebychev sur chacun des deux modèles introduits dans le Chapitre 1 ?

Pour le premier, rappelons que, pour tout numéro d'observation i ,

$$\forall \lambda > 0, \quad \mathbb{E}_\lambda[X_i] = \lambda^{-1}.$$

De fait, en utilisant le calcul effectué précédemment, il vient

$$\forall \lambda > 0, \forall a > 0, \quad \mathbb{P}_\lambda\{|X_i - \lambda^{-1}| \geq a\} \leq 2\lambda^{-2}a^{-2}.$$

Tournons-nous vers le second modèle. De façon similaire, pour tout numéro d'observation i

$$\forall p \in]0, 1[, \forall a > 0, \quad \mathbb{P}_p\{|X_i - p| \geq a\} \leq p(1-p)a^{-2}.$$

Comment comprendre ces deux estimations :

- (1) Dans le premier cas, nous comprenons que la probabilité d'observer une durée de vie loin de λ^{-1} est d'autant plus petite que λ est grand. Malheureusement, pour λ petit, l'estimation peut ne rien nous apporter : par exemple pour $a > \lambda$, l'estimation se réduit à une majoration par 2 d'une quantité inférieure à 1.
- (2) Le deuxième cas est un peu différent. Certes le membre de droite dépend à nouveau du paramètre, malheureusement inconnu, mais il est possible de **le majorer uniformément**.

En effet, simple étude de fonction montre que

$$(1.5) \quad \forall p \in]0, 1[, \quad p(1-p) \leq \frac{1}{4}.$$

A priori, nous sommes donc contraints à distinguer les deux cas, le second paraissant plus favorable. (Insistons : il serait stupide de conserver le paramètre d'intérêt dans le membre de droite. Il est en effet inconnu !)

1.3. Intervalles de confiance. Continuons l'étude du deuxième cas en injectant la borne (1.5). Nous obtenons

$$\forall p \in]0, 1[, \forall a > 0, \quad \mathbb{P}_p\{|X_i - p| \geq a\} \leq \frac{1}{4a^2}.$$

Bien que cette majoration soit moins bonne que la précédente (en raison du remplacement de $p(1-p)$ par $1/4$), nous comprenons combien elle est plus satisfaisante que la précédente du point de vue de l'estimation. La déviation a étant fixée, nous sommes en effet capables de majorer, indépendamment du paramètre, la probabilité que l'observation soit à distance supérieure à a du paramètre.

Par exemple, si $a = .7$, nous obtenons

$$\forall p \in]0, 1[, \quad \mathbb{P}_p\{|X_i - p| > .7\} \leq \frac{1}{4 * .7^2} \approx 0.51 \leq 0.52.$$

Nous en déduisons, qu'avec probabilité supérieure à 0.48, l'observation est à distance inférieure à 0.7 du paramètre. En réalité, l'observation étant inconnue et l'observation l'étant, nous préférons dire que le paramètre est à distance inférieure à 0.7 de l'observation avec probabilité supérieure à 0.48.

Par la suite, si x_i désigne l'observation effectuée pour la variable X_i , nous dirons que l'intervalle $]x_i - .7, x_i + .7[$ est un intervalle de confiance de niveau supérieur à .48 pour le paramètre p .

Naturellement, nous pouvons faire le niveau de confiance. Si ε désigne un réel dans $]0, 1[$ et si nous cherchons $a > 0$ tel que

$$\frac{1}{4a^2} = \varepsilon,$$

nous obtenons

$$a = \frac{1}{2\varepsilon^{1/2}}.$$

Avec le même vocabulaire que précédemment, nous en déduisons que l'intervalle $]x_i - \frac{1}{2\varepsilon^{1/2}}, x_i + \frac{1}{2\varepsilon^{1/2}}[$ est un intervalle de confiance pour le paramètre p de niveau de confiance supérieur à $1 - \varepsilon$.

Remarquons que cet intervalle peut ne pas être pertinent : par exemple, si $\varepsilon = .25$, alors $\varepsilon^{1/2} = .5$ et l'intervalle est de la forme $]x_i - 1, x_i + 1[$. Le résultat est de fait totalement inintéressant : puisque $]0, 1[$ est inclus dans $]x_i - 1, x_i + 1[$, nous savons déjà que le paramètre p est dans $]x_i - 1, x_i + 1[$. Nous comprendrons par la suite que l'estimation donnée ici est de mauvaise qualité car fondée sur une seule observation.

Néanmoins, nous pouvons donner la définition :

Définition. Soient $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ un modèle statistique, $g(\theta)$ une fonction (réelle) du paramètre θ à estimer, et S_1 et S_2 deux statistiques et α un réel entre 0 et 1.

Si s_1 et s_2 désignent les valeurs observées pour S_1 et S_2 , alors l'intervalle $]s_1, s_2[$ est appelé intervalle de confiance pour $g(\theta)$

de niveau de confiance égal (respectivement supérieur) à α si

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta\{S_1 < g(\theta) < S_2\} = (\text{resp. } \geq)\alpha.$$

1.4. Retour sur la première expérience. Comment obtenir un intervalle de confiance dans la première expérience? Il suffit par exemple de choisir a de la forme $a\lambda^{-1}$. Alors,

$$\forall \lambda > 0, \quad \mathbb{P}_\lambda\{|X_i - \lambda^{-1}| \geq a\lambda^{-1}\} \leq 2a^{-2}.$$

De fait,

$$\forall \lambda > 0, \quad \mathbb{P}_\lambda\{|\lambda X_i - 1| \geq a\} \leq 2a^{-2}.$$

Fixons maintenant un niveau α et cherchons a tel que $2a^{-2} = 1 - \alpha$. Il vient $a = \sqrt{2/(1 - \alpha)}$. Pour un tel choix de a , nous avons

$$\forall \lambda > 0, \quad \mathbb{P}_\lambda\{|\lambda X_i - 1| < a\} \geq \alpha,$$

i.e.

$$\forall \lambda > 0, \quad \mathbb{P}_\lambda\{|\lambda X_i - 1| < \sqrt{2/(1 - \alpha)}\} \geq \alpha.$$

En se rappelant que les valeurs de X_i sont toujours strictement positives, nous pouvons encore écrire

$$\forall \lambda > 0, \quad \mathbb{P}_\lambda \{ |\lambda - 1/X_i| < \sqrt{2/(1-\alpha)}/X_i \} \geq \alpha.$$

Ici, $S_1 = 1/X_i - \sqrt{2/(1-\alpha)}/X_i$ et $S_2 = 1/X_i + \sqrt{2/(1-\alpha)}/X_i$.

En fait, nous remarquons que $\sqrt{2/(1-\alpha)} \geq 1$ de sorte que la borne du dessous n'a pas d'intérêt. (Nous pouvons tout aussi bien choisir $S_1 = 0$.)

1.5. Introduction aux régions de confiance. Lorsque la fonction g du paramètre θ est à valeurs dans \mathbb{R}^d , avec $d \geq 2$, l'expression « intervalle de confiance » est remplacée par « région de confiance ». Voici un exemple :

Définition. Soient $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ un modèle statistique, $g(\theta)$ une fonction du paramètre θ à valeurs dans \mathbb{R}^2 à estimer, et S_1, S_2 et S_3 trois statistiques, S_3 à valeurs positives, et α un réel entre 0 et 1.

Si s_1, s_2 et s_3 désignent les valeurs observées pour S_1, S_2 et S_3 , alors la boule de centre (s_1, s_2) et de rayon s_3 est appelée boule de confiance pour $g(\theta)$ de niveau de confiance égal (respectivement supérieur) à α si

$$\forall \theta \in \Theta, \quad \mathbb{P}_\theta \{ \|g(\theta) - (S_1, S_2)\| < S_3 \} = (\text{resp. } \geq) \alpha.$$

(Ici, $\|\cdot\|$ désigne la norme euclidienne.)

2. Estimation empirique

Dans les deux exemples précédents, nous avons souligné la faible qualité des intervalles de confiance construits à partir d'une seule observation. Cet écueil est aisément compréhensible : la connaissance du modèle étudié ne peut être que limitée lorsque fondée sur une seule expérience ; naturellement, il est à espérer qu'elle se bonifie avec la répétition de l'expérience.

En statistiques, l'estimation porte le nom d'« empirique » lorsqu'appuyée sur la répétition d'une même expérience, dans des conditions similaires et indépendantes.

En réalité, l'utilisation d'une même expérience, mais répétée dans des conditions similaires et indépendantes, fait l'objet de plusieurs résultats de probabilité à connaître : loi des grands nombres et théorème central limite.

2.1. Loi des grands nombres. La loi des grands nombres traduit, sous forme mathématique, le principe rappelé ci-dessus : la répétition de l'expérience conduit à la connaissance.

Théorème. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X_1, \dots, X_n n variables aléatoires indépendantes de même loi, supposée intégrable et de carré intégrable. Alors, en désignant par m et par σ^2 sa moyenne et sa variance,

il vient

$$\forall a > 0, \quad \mathbb{P}\left\{\left|n^{-1} \sum_{i=1}^n X_i - m\right| \geq a\right\} \leq \frac{\sigma^2}{a^2 n}.$$

La preuve est une simple conséquence de l'inégalité de Bienaymé-Chebychev et du lemme suivant :

Lemme. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X_1, \dots, X_n n variables aléatoires indépendantes de même loi, supposée intégrable et de carré intégrable. Alors, la variance de la somme des X_1, \dots, X_n est égale à la somme des variances.

Notation. Usuellement, la quantité $n^{-1} \sum_{i=1}^n X_i$ est appelée moyenne empirique et notée \bar{X}_n .

Commentaire. Commentons le résultat : n désigne ici le nombre d'expériences considérées. Lorsque n tend vers l'infini, nous observons que le membre de droite tend vers 0. Autrement dit, la probabilité que la moyenne empirique \bar{X}_n soit éloignée de plus de a de la moyenne m tend vers 0 quand n tend vers l'infini : nous dirons que \bar{X}_n tend vers m en probabilité lorsque n tend vers l'infini.

En fait, nous pouvons même donner une estimation de la grandeur typique entre \bar{X}_n et m . Choisissons en effet a de la forme $a\sigma n^{-1/2}$, il vient

$$\forall a > 0, \quad \mathbb{P}\left\{|\bar{X}_n - m| \geq a n^{-1/2} \sigma\right\} \leq \frac{1}{a^2}.$$

Autrement dit, pour a de l'ordre de 10, la probabilité que la moyenne empirique et la vraie moyenne se trouve à une distance supérieure à $10n^{-1/2}\sigma$ est inférieure à .01.

Extensions. En réalité, il est possible d'étendre l'application de la loi des grands nombres à aux variables $f(X_1), \dots, f(X_n)$ pour une fonction f de \mathbb{R} dans lui-même (par exemple continue). En effet, l'indépendance et l'identique distribution sont conservées par composition par une même fonction : de fait, si f est intégrable et de carré intégrable sous la loi commune de X_1, \dots, X_n , nous pouvons affirmer

$$\forall a > 0, \quad \mathbb{P}\left\{\left|n^{-1} \sum_{i=1}^n f(X_i) - m_f\right| \geq a\right\} \leq \frac{\sigma_f^2}{a^2 n},$$

où m_f désigne $\mathbb{E}[f(X_1)]$ et σ_f^2 désigne $\mathbb{V}[f(X_1)]$.

2.2. Utilisation en statistiques. L'estimation empirique s'appuie sur la donnée d'un modèle échantillonné. Supposons de fait que $(\Omega, \mathcal{A}, \mu_\theta^{\otimes n})_{\theta \in \Theta}$ soit un modèle échantillonné de taille n et de lois $(\mu_\theta)_{\theta \in \Theta}$ muni de statistiques d'observation X_1, \dots, X_n .

Supposons que les observations soient intégrables et de carré intégrables. Nous posons

$$g(\theta) = \mathbb{E}_\theta[X_1] (= \mathbb{E}_\theta[X_i], \quad 1 \leq i \leq n).$$

Par définition, nous en déduisons que les statistiques X_1, \dots, X_n sont des estimateurs sans biais de $g(\theta)$. Par linéarité de l'espérance, nous en déduisons que la moyenne empirique

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

est aussi un estimateur sans biais de $g(\theta)$.

Par ailleurs, la loi des grands nombres montre que

$$\forall \theta \in \Theta, \forall a > 0, \quad \mathbb{P}_\theta\{|\bar{X}_n - g(\theta)| \geq a\} \leq \frac{\sigma_\theta^2}{na^2},$$

où $\sigma_\theta^2 = \mathbb{V}_\theta[X_1]$.

Plus généralement, si f désigne une fonction de \mathbb{R} dans \mathbb{R} (par exemple continue) et si les statistiques $f(X_1), \dots, f(X_n)$ sont intégrables et de carré intégrable, nous obtenons que

$$n^{-1} \sum_{i=1}^n f(X_i)$$

est un estimateur sans biais de $g_f(\theta) = \mathbb{E}_\theta[f(X_1)]$. De plus, la loi des grands nombres montre que

$$\forall \theta \in \Theta, \forall a > 0, \quad \mathbb{P}_\theta\left\{\left|n^{-1} \sum_{i=1}^n f(X_i) - g_f(\theta)\right| \geq a\right\} \leq \frac{\mathbb{V}_\theta(f(X_1))}{na^2}.$$

Exemple d'utilisation. Reprenons l'exemple numéro 2 du contrôle qualité. En reprenant les calculs effectués dans la première partie, nous obtenons que

$$\forall p \in]0, 1[, \forall a > 0, \quad \mathbb{P}_p\{|\bar{X}_n - p| \geq a\} \leq \frac{p(1-p)}{na^2}.$$

En majorant $p(1-p)$ par $1/4$ comme nous l'avons déjà fait ci-dessus, nous en déduisons que

$$\forall p \in]0, 1[, \forall a > 0, \quad \mathbb{P}_p\{|\bar{X}_n - p| \geq a\} \leq \frac{1}{4na^2}.$$

En remplaçant a par $(2n^{1/2})^{-1}a$, il vient finalement

$$\forall p \in]0, 1[, \forall a > 0, \quad \mathbb{P}_p\{|\bar{X}_n - p| \geq \frac{a}{2n^{1/2}}\} \leq \frac{1}{a^2}.$$

Nous en déduisons : si (x_1, \dots, x_n) désignent les réalisations des observations dans l'exemple 2 du chapitre 1 et si $\bar{x}_n = (x_1 + \dots + x_n)/n$ désigne la

réalisation de la moyenne empirique, alors l'intervalle

$$]\bar{x}_n - \frac{a}{2n^{1/2}}, \bar{x}_n + \frac{a}{2n^{1/2}}[$$

est un intervalle de confiance de niveau supérieur à $1 - a^{-2}$. Par exemple, pour $n = 1000$ (valeur numérique choisie dans le chapitre 1) et $a = 10$, l'intervalle de confiance est de largeur $10 \times 1000^{-1/2} \approx 0.31$ et de niveau supérieur à .99.

Insistons sur la significativité de ce résultat : la fourchette de l'intervalle de confiance est régie par le nombre d'observations et, plus précisément, par la racine du nombre d'observations.

2.3. Théorème Limite Central. Une question naturelle est de savoir si la fourchette donnée par l'inégalité de Bienaymé-Chebychev est optimale ou non.

La réponse à cette question est donnée par le théorème limite central. Le mot « central » a ici la signification suivante : le théorème a un champ d'application très large, pour ne pas dire universel, qui lui donne une place centrale, au coeur de la théorie des probabilités.

Théorème. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même loi intégrable et de carré intégrable. En désignant par $m = \mathbb{E}(X)$ et $\sigma^2 = \mathbb{V}(X)$ et en supposant $\sigma^2 > 0$, il vient, pour tout a, b :

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left\{a < n^{1/2} \frac{\bar{X}_n - m}{\sigma} < b\right\} = \int_a^b \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}}.$$

Nous ne démontrerons pas ce résultat. Faisons en revanche les remarques suivantes :

- (1) Le terme de droite correspond à la probabilité qu'une variable aléatoire de loi gaussienne centrée réduite appartienne à l'intervalle $]a, b[$ (ou $[a, b]$), i.e.

$$\int_a^b \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} = \mathbb{P}\{a < Z < b\},$$

où Z est une variable aléatoire de loi de densité

$$x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

- (2) Le résultat se focalise sur les *fluctuations* de la moyenne empirique : asymptotiquement, elles se comportent de façon gaussienne. Ceci explique le rôle central des gaussiennes en théorie des probabilités.

- (3) Le membre de gauche peut être réécrit sous une forme plus proche de celle attendue pour l'estimation :

$$\lim_{n \rightarrow +\infty} \mathbb{P}\{\bar{X}_n - a\sigma n^{-1/2} < m < \bar{X}_n - b\sigma n^{-1/2}\} = \int_a^b \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}}.$$

En réalité, un simple changement de variable dans l'intégrale de droite permet de remplacer a par $-a$ et b par $-b$, i.e.

$$\lim_{n \rightarrow +\infty} \mathbb{P}\{\bar{X}_n + a\sigma n^{-1/2} < m < \bar{X}_n + b\sigma n^{-1/2}\} = \int_a^b \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}}.$$

- (4) Soulignons enfin que la loi gaussienne est tabulée (par exemple sur ordinateur, instruction `cdfnor` en Scilab) : cela signifie que l'on peut calculer une approximation de l'intégrale ci-dessus (dont aucune formule explicite n'est en revanche connue). Par exemple, si $b = -a = 1.960$, il vient

$$\int_{-1.960}^{1.960} \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \approx 0.950 \quad (\geq).$$

De façon comparable, si $b = -a = 2.576$, il vient

$$\int_{-2.576}^{2.576} \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \approx 0.990 \quad (\geq).$$

Tournons maintenant vers l'interprétation pratique de ce résultat. Il doit être compris de la façon suivante : essentiellement, lorsque n est grand, la probabilité que la moyenne soit à une distance inférieure à $1.960 \times \sigma \times n^{-1/2}$ de la moyenne empirique observée est (presque) égale à 0.95. De façon similaire, la probabilité que la moyenne soit à une distance inférieure à $2.576 \times \sigma \times n^{-1/2}$ de la moyenne empirique observée est (presque) égale à 0.99. Si nous comparons ce dernier résultat avec celui obtenu par application de l'inégalité de Bienaymé-Chebychev, nous comprenons deux choses :

- (1) La longueur de la fourchette est bien proportionnelle à $n^{-1/2}$.
- (2) En revanche, le coefficient de proportionnalité est plus petit : le théorème limite central fournit une égalité à la limite alors que l'inégalité de Bienaymé-Chebychev ne donne qu'une majoration. Par exemple, l'inégalité de Bienaymé-Chebychev donne une fourchette de rayon $10 \times \sigma \times n^{-1/2}$ au niveau 0.99 contre $2.576 \times \sigma \times n^{-1/2}$ pour le théorème central limite.

Soulignons pour conclure ce paragraphe que le théorème central limite s'applique aussi à une suite de la forme $(f(X_n))_{n \geq 1}$, les variables $(X_n)_{n \geq 1}$ étant indépendantes et identiquement distribuées, pour peu que $f(X_1)$ soit de carré intégrable et de variance non nulle. L'application se fait sur le modèle de l'application de la loi des grands nombres.

2.4. Intervalles de confiance asymptotiques. Nous comprenons que le théorème central limite fournit une sorte d'intervalle de confiance, mais valable simplement pour un nombre infini (ou très grand) d'observations. Ceci justifie la terminologie suivante :

Définition. Soient $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ un modèle statistique, $g(\theta)$ une fonction réelle du paramètre θ à estimer, $(S_n^1)_{n \geq 1}$ et $(S_n^2)_{n \geq 1}$ deux suites de statistiques et α un réel entre 0 et 1.

Si $(s_n^1)_{n \geq 1}$ et $(s_n^2)_{n \geq 1}$ désignent les valeurs observées pour $(S_n^1)_{n \geq 1}$ et $(S_n^2)_{n \geq 1}$, alors la suite d'intervalles $(]s_n^1, s_n^2[)_{n \geq 1}$ est appelée intervalle de confiance asymptotique pour $g(\theta)$ de niveau de confiance égal (respectivement supérieur) à α si

$$\forall \theta \in \Theta, \quad \lim_{n \rightarrow +\infty} \mathbb{P}\{S_n^1 \leq g(\theta) \leq S_n^2\} = (\text{resp. } \geq) \alpha.$$

Application. Revenons à l'exemple du contrôle qualité. Ici, la variance des observations vaut $p(1-p)$ sous chaque \mathbb{P}_p , $p \in]0, 1[$. (Nous conservons les mêmes notations que précédemment sans les rappeler.) Une application directe du théorème limite central montre alors que

$$\forall p \in]0, 1[, \quad \lim_{n \rightarrow +\infty} \mathbb{P}_p\{\bar{X}_n - 1.960[p(1-p)]^{1/2}n^{-1/2} < p < \bar{X}_n + 1.960[p(1-p)]^{1/2}n^{-1/2}\} \approx 0.95.$$

Nous constatons le même problème que celui rencontré pour lors de l'estimation à nombre d'observations fini : la variance dépend du paramètre à estimer et est, de fait, inconnue.

Pour contourner cette difficulté, il est possible de reproduire la même stratégie que celle mise en place précédemment : nous pouvons toujours majorer $p(1-p)$ par $1/4$. De sorte que

$$\begin{aligned} & \mathbb{P}_p\{\bar{X}_n - 1.960[p(1-p)]^{1/2}n^{-1/2} < p < \bar{X}_n + 1.960[p(1-p)]^{1/2}n^{-1/2}\} \\ & \leq \mathbb{P}_p\{\bar{X}_n - 1.960(2n)^{-1/2} < p < \bar{X}_n + 1.960(2n)^{-1/2}\}. \end{aligned}$$

Lorsque n tend vers l'infini, rien ne dit que le terme de droite ci-dessus ait une limite. En revanche, nous comprenons bien qu'il vit au-dessus de la limite du terme de gauche. De fait, en désignant par $(\bar{x}_n)_{n \geq 1}$ une suite de réalisations de la moyenne empirique, nous pouvons écrire que la suite d'intervalles $(] \bar{x}_n - 1.960(2n)^{-1/2}, \bar{x}_n + 1.960(2n)^{-1/2} [)_{n \geq 1}$ est un intervalle de confiance asymptotique de niveau de confiance supérieur à .95 pour le paramètre p .

2.5. Estimation de la variance. Une autre méthode, dans l'application ci-dessus, pour contourner la méconnaissance de la variance consiste à la remplacer par une estimation fabriquée à l'aide des observations.

Définition. Etant donné un modèle statistiques $(\Omega, \mathcal{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$ muni de n ,

$n \geq 2$, statistiques X_1, \dots, X_n indépendantes et de même loi μ_θ sous chaque \mathbb{P}_θ , $\theta \in \Theta$, nous appelons variance empirique non-biaisée la quantité :

$$\bar{\Sigma}_n = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Comprenons quelle est le lien entre l'estimateur $\bar{\Sigma}_n$ et la fonction $\theta \in \Theta \mapsto \mathbb{V}_\theta(X_1)$.

Proposition. Sous les notations de la définition ci-dessus, la variance empirique non-biaisée est un estimateur sans biais de la fonction $\theta \in \Theta \mapsto \mathbb{V}_\theta(X_1)$ dès lors que les lois $(\mu_\theta)_\theta$ sont de carré intégrable, i.e.

$$\forall \theta \in \Theta, \quad \mathbb{E}_\theta[\bar{\Sigma}_n] = \mathbb{V}_\theta(X_1).$$

Preuve. Nous commençons par développer le carré dans la définition de la variance empirique non-biaisée :

$$\begin{aligned} \sum_{k=1}^n (X_k - \bar{X}_n)^2 &= \sum_{k=1}^n X_k^2 + n\bar{X}_n^2 - 2\bar{X}_n \sum_{k=1}^n X_k \\ &= \sum_{k=1}^n X_k^2 - n\bar{X}_n^2 \\ &= \sum_{k=1}^n X_k^2 - n^{-1} \left(\sum_{k=1}^n X_k^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j \right) \\ &= \frac{n-1}{n} \sum_{k=1}^n X_k^2 - 2n^{-1} \sum_{1 \leq i < j \leq n} X_i X_j. \end{aligned}$$

Prenons maintenant l'espérance de chaque côté :

$$\mathbb{E}_\theta \left[\sum_{k=1}^n (X_k - \bar{X}_n)^2 \right] = \frac{n-1}{n} \sum_{k=1}^n \mathbb{E}_\theta X_k^2 - 2n^{-1} \sum_{1 \leq i < j \leq n} \mathbb{E}_\theta [X_i X_j].$$

Maintenant, en utilisant l'indépendance et l'identique distribution des statistiques, nous obtenons :

$$\begin{aligned} \mathbb{E}_\theta \left[\sum_{k=1}^n (X_k - \bar{X}_n)^2 \right] &= (n-1) \mathbb{E}_\theta(X_1^2) - 2n^{-1} \sum_{1 \leq i < j \leq n} \mathbb{E}_\theta^2(X_1) \\ &= (n-1) [\mathbb{E}_\theta(X_1^2) - \mathbb{E}_\theta^2(X_1)]. \end{aligned}$$

Le résultat découle.

Une deuxième question se pose maintenant : quand n devient grand, la variance empirique ressemble-t-elle à la vraie variance. Pour répondre, à cette

question, nous allons supposer que, pour chaque valeur de $\theta \in \Theta$, $\mathbb{E}_\theta(X_k^4)$ existe pour $1 \leq k \leq n$. Alors, la loi des grands nombres garantit que

$$\forall \theta \in \Theta, \forall a > 0, \quad \mathbb{P}_\theta \left\{ \left| n^{-1} \sum_{k=1}^n X_k^2 - \mathbb{E}_\theta(X_1^2) \right| > a \right\} \leq \frac{\mathbb{V}_\theta(X_1^2)}{a^2 n}.$$

De sorte que, pour chaque θ , avec grande probabilité sous \mathbb{P}_θ ,

$$\frac{1}{n} \sum_{k=1}^n X_k^2 \approx \mathbb{E}_\theta(X_1^2)$$

lorsque n est grand. En fait, nous pouvons remplacer $1/n$ par $1/(n-1)$ dans l'approximation ci-dessus, i.e.

$$\frac{1}{n-1} \sum_{k=1}^n X_k^2 \approx \mathbb{E}_\theta(X_1^2)$$

Par ailleurs, nous savons, qu'avec grande probabilité sous \mathbb{P}_θ , pour n grand,

$$\bar{X}_n \approx \mathbb{E}_\theta(X_1).$$

Donc,

$$\frac{n}{n-1} \bar{X}_n^2 \approx \mathbb{E}_\theta^2(X_1).$$

A l'arrivée, nous pouvons affirmer que la variance empirique ressemble, avec grande probabilité sous \mathbb{P}_θ à $\mathbb{V}_\theta(X_1)$, pour chaque $\theta \in \Theta$.

En réalité, ce raisonnement peut être mené de façon tout-à-fait rigoureuse.

En particulier, nous admettrons

Théorème. Soit $(\Omega, \mu_\theta^{\otimes n})_{\theta \in \Theta}$ un modèle échantillonné de taille n muni de n statistiques X_1, \dots, X_n de carré intégrable. (Nous supposons que la variance de X_1 sous chaque \mathbb{P}_θ est non nulle.) Alors, en désignant par \bar{X}_n la moyenne empirique et par $\bar{\Sigma}_n$ la variance empirique non-biaisée, il vient, pour tout a, b :

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta \{ a \bar{\Sigma}_n < n^{1/2} [\bar{X}_n - m] < b \bar{\Sigma}_n \} = \int_a^b \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}}.$$

En particulier, en désignant par \bar{x}_n et $\bar{\sigma}_n$ les réalisations de la moyenne empirique et de la variance empirique, nous en déduisons que les intervalles $(\lfloor \bar{x}_n + a \bar{\sigma}_n n^{-1/2}, \bar{x}_n + b \bar{\sigma}_n n^{-1/2} \rfloor)_{n \geq 1}$ forment un intervalle de confiance asymptotique pour $g(\theta) = \mathbb{E}_\theta(X_1)$ de niveau $\mathbb{P}\{a < Z < b\}$, où Z suit une loi gaussienne centrée réduite.

En particulier, en choisissant $b = -a = 1.960$, il vient $(\lfloor \bar{x}_n - 1.960 \bar{\sigma}_n n^{-1/2}, \bar{x}_n + 1.960 \bar{\sigma}_n n^{-1/2} \rfloor)_{n \geq 1}$ intervalle de confiance asymptotique de niveau de confiance 0.95.

En pratique, on remplace n tend vers l'infini par n grand. (On retient n supérieur au millier de façon générale.)

Exemple. On interroge 1000 individus au hasard dans la population pour connaître leur opinion sur un projet d'aménagement public. Une proportion $p = .55$ se déclarent favorables avec une variance empirique de 0.24. En modélisant le sondage par une loi binomiale et en acceptant $n = 1000$ comme suffisamment grand, on en déduit comme intervalle de confiance asymptotique de niveau de confiance 0.95 l'intervalle $]0.535, .565[$ pour la vraie proportion d'individus favorables au projet d'aménagement public.

2.6. Résumé. Pour n grand, on utilisera le TLC. Pour n petit, on utilisera l'inégalité de Bienaymé-Chebychev.

Chapitre 4. Modèles Gaussiens.

Le cas gaussien occupe, aussi bien en probabilités qu'en statistiques, une position particulière : en raison du théorème limite central, la somme de petits aléas indépendants et identiquement distribués conduit asymptotiquement à une répartition ou encore à une distribution gaussienne du hasard.

Ceci explique pourquoi de nombreux phénomènes dont les fluctuations s'expliquent par l'accumulation de petits facteurs indépendants suivent un régime gaussien. Il en va aussi bien de la répartition de la taille ou du poids des individus dans la population que des fluctuations de la poussée du moteur d'une fusée au moment de son décollage.

Ce Chapitre a pour objectif de rappeler les principes essentiels des variables gaussiennes et d'en expliquer l'utilisation en statistiques, particulièrement en estimation.

1. Rappels : Distribution Gaussienne

1.1. Densité Gaussienne Centrée Réduite. Nous commençons par rappeler :

DÉFINITION 1.1. *On appelle densité gaussienne centrée réduite la fonction*

$$g : x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

En particulier, nous admettrons que l'intégrale sur \mathbb{R} tout entier de la fonction g est (finie et) égale à 1.

(Faire un dessin : $g(1) = g(-1) = g'(-1) = -g(1) \approx 0,24$, $g(0) \approx 0,40$. Ecrire les quantiles.)

La fonction g étant une densité, nous pouvons lui associer des variables aléatoires :

DÉFINITION 1.2. *Etant donné un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, une variable aléatoire X est dite de loi gaussienne centrée réduite si sa loi admet g pour densité, i.e.*

$$\mathbb{P}\{a \leq X \leq b\} = \int_a^b g(x) dx, \quad a < b.$$

Attention : on ne connaît pas de primitive explicite pour g !

La terminologie centrée réduite s'explique par la remarque suivante :

PROPOSITION 1.3. *Sous les hypothèses et notations de la définition ci-dessus, la variable X est de carré intégrable et*

$$\mathbb{E}(X) = 0 \quad \text{et} \quad \mathbb{V}(X) = \mathbb{E}(X^2) = 1.$$

Insistons : le mot « centrée » signifie de moyenne nulle et le mot « réduite » de moyenne 1.

Preuve. L'intégrabilité à l'ordre 2 est une conséquence de la forme g : la présence d'une exponentielle (qui plus est d'une exponentielle d'un carré) assure l'intégrabilité à l'ordre deux.

Le calcul de l'espérance est immédiat. La densité est paire. Donc, la fonction $x \in \mathbb{R} \mapsto xg(x)$ est impaire (et intégrable). Son intégrale sur \mathbb{R} est nulle.

Examinons maintenant le calcul de la variance. Clairement, par parité, nous pouvons ramener l'intégrale sur $[0, +\infty)$:

$$\mathbb{E}[X^2] = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx.$$

Il reste à effectuer une intégration par parties avec $u(x) = x$ et $v'(x) = x \exp(-x^2/2)$. Clairement $u'(x) = 1$ et $v(x) = -\exp(-x^2/2)$. De fait, nous sommes ramener au calcul de l'intégrale de g elle-même. \square

1.2. Modification de la moyenne et de la variance. Une façon de modifier la moyenne et la variance consiste à multiplier et à translater une variable de loi gaussienne.

PROPOSITION 1.4. *Etant donné un espace de probabilité muni d'une v.a. X de loi gaussienne centrée réduite et $m \in \mathbb{R}$ et $\sigma > 0$ deux réels, la v.a. $m + \sigma X$ admet m pour moyenne et σ^2 pour variance. De plus, sa loi admet une densité de la forme*

$$g_{m,\sigma} : x \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

La preuve est obtenue par simple changement de variable. On dira que $m + \sigma X$ suit une loi gaussienne de moyenne m et de variance σ^2 et on notera $m + \sigma X \sim \mathcal{N}(m, \sigma^2)$.

Remarque 1. Réciproquement, si Y suit une loi $\mathcal{N}(m, \sigma^2)$, alors $\sigma^{-1}(Y - m)$ suit une loi gaussienne centrée réduite.

Remarque 2. Lorsque $\sigma = 0$, $m + \sigma X = m$! La variable est constante. On dira aussi qu'il s'agit d'une variable gaussienne, mais constante.

(Faire un dessin en faisant la moyenne et la variance. Placer les quantiles)

1.3. Somme de variables aléatoires gaussiennes indépendantes.

Nous admettrons le résultat suivant :

PROPOSITION 1.5. *Etant donné un espace de probabilité munis de deux variables aléatoires gaussiennes INDEPENDANTES X et Y de lois gaussiennes $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$, la variable aléatoire $X + Y$ suit une loi gaussienne $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Notons bien que l'additivité des moyennes et des variances découle de résultats précédemment cités. La nouveauté réside ici dans la stabilité du caractère gaussien.

2. Construction d'intervalles de confiance pour la moyenne

Considérons le problème suivant. Une usine réceptionne une nouvelle machine-outil capable de découper des barres d'acier. La longueur des tiges est ajustable par l'opérateur à l'aide d'une molette. Le constructeur garantit qu'avec probabilité 0,95, les tiges produites sont, quelle que soit la longueur de production, découpées avec une marge d'erreur inférieure à $1,960 \times 5$ mm.

Pour calibrer le réglage de la machine, il est néanmoins conseillé de vérifier les graduations de la molette en estimant la longueur des tiges produites lorsque la molette est en position « 1 m ».

2.1. Modèle statistique gaussien. Il faut ici savoir déchiffrer l'énoncé : nous pouvons clairement assimiler la longueur des tiges produites à une variable de loi gaussienne. Manifestement, la longueur moyenne n'est pas explicitement connue lorsque la molette est en position 1m. En revanche, l'écart type est fixé égal à $5\text{mm} = .5$ cm.

Si jamais n tiges sont produites en guise d'essai, nous sommes amenés à considérer un modèle statistique de la forme

$$(\Omega, \mathbb{P}_m = \mathcal{N}(m, .5)^{\otimes n})_{m \in \mathbb{R}}.$$

Naturellement, nous pourrions restreindre l'espace du paramètre m : nous comprenons bien que m est attendu à être positif. Ici, cela n'a pas beaucoup d'importance.

Nous considérons de fait n statistiques X_1, \dots, X_n indépendantes et de loi $\mathcal{N}(m, .5)$ sous chaque \mathbb{P}_m .

Nous sommes alors intéressés pour mettre en place une procédure d'estimation du paramètre m .

2.2. Estimateur Empirique à variance connue. La construction d'un intervalle de confiance s'appuie sur le résultat suivant, à bien connaître :

PROPOSITION 2.1. *Sous les hypothèses ci-dessus, sous chaque \mathbb{P}_m , la moyenne empirique suit une loi gaussienne $\mathcal{N}(m, 0.5^2/n)$. En particulier,*

$$\forall m \in \mathbb{R}, \quad \mathbb{P}\left\{a < \frac{\bar{X}_n - m}{0.5^2/n} < b\right\} = \mathbb{P}\{a < Z < b\},$$

où Z suit une loi gaussienne centrée réduite.

Comme Corollaire, nous en déduisons : si \bar{x}_n désigne la valeur observée pour la moyenne empirique, alors $]\bar{x}_n - 1.960\sqrt{.5}\sqrt{n}, \bar{x}_n + 1.960\sqrt{.5}\sqrt{n}[$ est un intervalle de confiance pour m de niveau de confiance 0,95.

Attention, cet intervalle de confiance est non-asymptotique ! Il s'agit là de l'intérêt du cas gaussien : néanmoins, sa construction est complètement similaire à celle envisagée à l'aide du théorème central limite dans le cas d'un nombre "infini" d'observations.

2.3. Estimation de la moyenne à variance inconnue. En pratique, il est en réalité rare de connaître précisément la variance expliquant les observations. Il est de fait nécessaire de savoir construire un intervalle de confiance pour la moyenne indépendamment de la connaissance de la variance.

En repensant à la construction des intervalles de confiance asymptotiques à l'aide du théorème central limite, nous comprenons que la variance empirique peut avoir, dans cette perspective, un rôle essentiel.

Nous distinguerons par conséquent deux cas :

- (1) Si le nombre d'observations est grand, il est possible de remplacer "brutalement" la variance par la variance empirique et de proposer un intervalle de confiance asymptotique.
- (2) Si le nombre d'observations est petit, le caractère asymptotique ne s'applique pas : il est nécessaire d'approfondir l'étude.

Une première question est de savoir quelle est la forme du modèle statistique. Pour indiquer la méconnaissance de la variance, il est nécessaire de la mentionner comme un paramètre :

$$(\Omega, \mathcal{N}(m, \sigma)^{\otimes n})_{m \in \mathbb{R}, \sigma > 0},$$

n désignant naturellement le nombre d'observations. Naturellement, le modèle est muni des statistiques d'observations : X_1, \dots, X_n .

L'estimateur de la moyenne demeure le même : il s'agit de considérer la moyenne empirique \bar{X}_n . Afin de contrôler la distance entre \bar{X}_n et m , nous proposons de considérer le rapport

$$\sqrt{n} \frac{\bar{X}_n - m}{\bar{\Sigma}_n},$$

où $\bar{\text{Sigma}}_n^2$ désigne la variance empirique non biaisée, i.e.

$$\bar{\Sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad n \geq 2.$$

Avant de poursuivre, comprenons bien que la statistique

$$T_n = \sqrt{n} \frac{\bar{X}_n - m}{\bar{\Sigma}_n}$$

se comporte, pour n grand, comme une gaussienne centrée réduite. Il s'agit là d'une conséquence du théorème central limite. Autrement dit, nous sommes capables de dire que l'intervalle $]\bar{x}_n - 1,960\bar{\sigma}_n n^{-1/2}, \bar{x}_n + 1,960\bar{\sigma}_n n^{-1/2}[$ est un intervalle de confiance asymptotique de niveau 0,95 si \bar{x}_n et $\bar{\sigma}_n$ désignent les valeurs expérimentales de la moyenne empirique et de la variance empirique.

En revanche, nous insistons : l'intervalle de confiance n'est pas valable si n est petit.

L'idée est la suivante :

PROPOSITION 2.2. *Etant donné le modèle statistique*

$$(\Omega, \mathcal{N}(m, \sigma)^{\otimes n})_{m \in \mathbb{R}, \sigma > 0},$$

muni de n observations X_1, \dots, X_n indépendantes et de même loi $\mathcal{N}(m, \sigma^2)$ sous chaque $\mathbb{P}_{m, \sigma}$, la loi de la statistique

$$T_n = \sqrt{n} \frac{\bar{X}_n - m}{\bar{\Sigma}_n}$$

ne dépend ni de m ni de σ^2 et est appelée loi de Student de paramètre $n - 1$.

Nous admettrons ce résultat.

Expliquons en l'intérêt : si la loi de Student de degré $n - 1$ est tabulée, i.e. que nous connaissons t_n^α , $\alpha \in]0, 1[$, tel que

$$\mathbb{P}\{-t_n^\alpha \leq \mathcal{T}_n \leq t_n^\alpha\} = 1 - 2\alpha,$$

pour \mathcal{T}_n variable aléatoire de loi de Student de degré $n - 1$, alors nous sommes capables de dire avec les notations précédentes que

$$\mathbb{P}\{-t_n^\alpha \leq T_n \leq t_n^\alpha\} = 2\alpha - 1,$$

de sorte que l'intervalle $]\bar{x}_n - t_n^\alpha \bar{\sigma}_n n^{-1/2}, \bar{x}_n + t_n^\alpha \bar{\sigma}_n n^{-1/2}[$ est un intervalle de confiance de niveau $2\alpha - 1$ pour m . (Ici, \bar{x}_n et $\bar{\sigma}_n$ désignent les réalisations de la moyenne et de l'écart-type empiriques.)

2.4. Loi de Student de degré n . Voici en réalité la définition de la loi de Student de degré n :

DÉFINITION 2.3. *La loi de Student de degré n est la loi de densité :*

$$x \in \mathbb{R} \mapsto \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Ici,

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} \exp(-x) dx.$$

Nous remarquons que la densité de Student de degré n est paire. Par ailleurs, lorsque n tend vers l'infini, pour t fixé,

$$\begin{aligned} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} &= \exp\left[-\frac{n+1}{2} \ln\left(1 + \frac{t^2}{n}\right)\right] \\ &\approx \exp\left[-\frac{n+1}{n} \frac{t^2}{2}\right]. \\ &\rightarrow \exp\left(-\frac{t^2}{2}\right). \end{aligned}$$

Nous comprenons que la partie variable de la densité tend, avec n , vers la partie variable de la densité gaussienne. En fait, nous pourrions montrer qu'il en va de même de la partie constante : ceci explique pourquoi, pour n grand, il est possible de ramener le problème au cas gaussien, comme nous l'avons déjà expliqué.

La moyenne est nulle et la variance vaut $n/(n-2)$ pour $n \geq 3$.

En pratique, la loi de Student est tabulée, i.e. que les valeurs des t_n^α sont données dans une table :

n	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

2.5. Application. Avant d'être mis sur le marché, un nouveau médicament est testé sur un groupe de sujets sains afin d'évaluer son impact sur le taux d'hémoglobine dans le sang (l'hémoglobine permet le transport de l'oxygène).

Vu le prix très élevé du médicament, il est décidé de procéder à un premier test sur un échantillon restreint de 30 personnes.

La moyenne empirique relevée est de 15,7 grammes par déciLitre avec un écart-type empirique de 2,6. Donner un intervalle de confiance pour la moyenne gouvernant le modèle en retenant une hypothèse gaussienne.

Réponse : $]15.7 - 2.045 \times 2.6 \times 30^{-1/2}, 15.7 + 2.045 \times 2.6 \times 30^{-1/2}[$ au niveau 0.95, soit $]14, 7; 16, 7[$.

3. Construction d'intervalle de confiance pour l'écart-type

Dans la section précédente, nous avons vu comment estimer la moyenne dans un modèle gaussien, avec ou sans connaissance de la variance.

Nous nous tournons maintenant vers l'estimation de l'écart-type.

Donnons un exemple : supposons que la taille en cm d'une femme soit modélisée par une loi normale de moyenne m et d'écart-type σ . Afin de prévoir la production de la collection hiver 2010-2011, une chaîne de prêt à porter doit avoir une idée de la taille moyenne des femmes, mais aussi de l'écart-type régissant la distribution : cela lui donnera une idée de la proportion de pantalons en taille 34, 36, 38, 40 à produire.

3.1. Principe. Le principe s'appuie sur la proposition suivante :

PROPOSITION 3.1. *Etant donné le modèle statistique*

$$(\Omega, \mathcal{N}(m, \sigma)^{\otimes n})_{m \in \mathbb{R}, \sigma > 0},$$

muni de n observations X_1, \dots, X_n indépendantes et de même loi $\mathcal{N}(m, \sigma^2)$ sous chaque $\mathbb{P}_{m, \sigma}$, la loi de la statistique

$$U_n = (n - 1) \bar{\Sigma}_n^2 \sigma^2$$

ne dépend ni de m ni de σ et est appelée loi du ξ^2 de degré $n - 1$, notée $\chi^2(n - 1)$.

En particulier, si la loi du χ^2 de degré $n - 1$ est tabulée, i.e. que nous connaissons q_n^α , $\alpha \in]0, 1[$, tel que

$$\mathbb{P}\{q_n^\beta \leq \mathcal{X}_n \leq q_n^\alpha\} = \alpha - \beta,$$

pour \mathcal{X}_n variable aléatoire de loi du χ^2 de degré $n - 1$, alors nous sommes capables de dire avec les notations précédentes que

$$\mathbb{P}\{q_n^\beta \leq U_n \leq q_n^\alpha\} = \alpha - \beta,$$

de sorte que l'intervalle $]\sqrt{n-1}\bar{\sigma}_n/q_n^\alpha, \sqrt{n-1}\bar{\sigma}_n/q_n^\beta[$ est un intervalle de confiance de niveau $\alpha - \beta$ pour σ . (Ici, $\bar{\sigma}_n$ désignent la réalisation de l'écart-type empirique.)

3.2. Loi du chi-deux de degré n . Voici en réalité la définition de la loi du chi-deux de degré n :

DÉFINITION 3.2. *La loi du chi-deux de degré n est la loi de densité :*

$$t \in \mathbb{R} \mapsto \frac{1}{2^{k/2}\Gamma(k/2)} t^{k/2-1} \exp(-t/2) \mathbf{1}_{t>0}.$$

Sa moyenne vaut n et sa variance $2n$.

Voici la table associée

n	0,005	0.010	0.025	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.975	0.990	0.995
1	0.0000	0.0002	0.0010	0.0039	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.879
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.591
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	13.277
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	15.507
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	17.535
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	19.678
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	21.024
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.705
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	24.433
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	26.216
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	28.053
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	29.945
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	31.881
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	33.861
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	35.885
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	37.953
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	32.0	33.4	39.965
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	41.921
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	43.821
20	7.43	8.26	9.56	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	45.665
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	47.453
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	49.185
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	50.871
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	52.511
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	54.105
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.1	35.6	38.9	41.9	45.6	55.653
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	57.155
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	58.611
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	60.021
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	61.385
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	71.421
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	81.335
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.701
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	105.411
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.4	119.651
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	133.511
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	144.211

CHAPITRE 5

Chapitre 5. Tests.

1. Principes

1.1. Objectif. Etant donné un modèle statistique $(\Omega, \mathbb{P}_\theta)_{\theta \in \Theta}$ modélisant un phénomène observable et une partition de l'ensemble des paramètres sous la forme $\Theta = \Theta_0 \cup \Theta_1$, un test d'hypothèses vise à déterminer si, au vu des observations expérimentales, il est raisonnable d'accepter le paramètre θ expliquant l'expérience comme un élément de Θ_0 ou de Θ_1 .

1.2. Premier exemple : notions d'hypothèses. Reprenons l'exemple du chapitre précédent : une machine outil découpe des tiges d'acier avec un écart type de 5 mm ; après un arrêt prolongé, un ouvrier la remet en route et décide, avant de l'utiliser à des fins industrielles, de vérifier si elle est correctement ajustée à la découpe de tiges de 1m en la testant sur 10 tiges successives. Voici les résultats qu'il obtient (en mm) :

988	997	995	989	997	985	1000	995	991	999
-----	-----	-----	-----	-----	-----	------	-----	-----	-----

Au vu de ces observations, peut-on considérer que la machine est dérégulée ?

Afin de répondre à cette question, il est nécessaire de commencer par la formaliser mathématiquement. Pour cela, nous modélisons la taille de découpe par une variable gaussienne de paramètre μ inconnu (en mm) et d'écart type 5. Autrement dit, nous sommes amenés à considérer le modèle gaussien :

$$(\Omega, \mathcal{N}(\mu, \sigma^2)^{\otimes n})_{\mu \in \mathbb{R}},$$

avec $n = 10$ et $\sigma = 5$.

Nous comprenons de fait que l'ensemble des paramètres peut-être partitionné en deux :

$$\Theta_0 = \{\mu : \mu = 1000\}, \quad \Theta_1 = \{(\mu, \sigma) : \mu \neq 1000\}.$$

Il s'agit de décider, au vu des observations, entre deux hypothèses :

- (1) H_0 : “ $\mu \in \Theta_0$ ” i.e. “la machine n'est pas dérégulée”,
- (2) H_1 : “ $\mu \in \Theta_1$ ” i.e. “la machine est dérégulée”.

Vocabulaire. L'hypothèse H_0 porte le nom d'hypothèse nulle : elle correspond, d'un point de vue pratique, à un état initial. Il s'agit de fait de décider s'il est préférable de la conserver ou non au regard des observations

expérimentales. L'hypothèse H_1 porte le nom d'hypothèse alternative : d'un point de vue pratique, la vérification de l'hypothèse alternative est supposée conduire à des modifications sensibles du réel. Dans le contexte courant, nous comprenons que la détection d'un biais dans la découpe conduirait à un nouveau réglage de la machine.

1.3. Risques. Naturellement, l'aide à la décision étant formulée à partir de l'observation, il existe un risque de rendre une mauvaise décision. Ce risque est en réalité double :

- (1) Il existe le risque de décider d'adopter l'hypothèse alternative à tort, i.e. de décider H_1 alors que H_0 est vraie.
- (2) Il existe le risque de décider de conserver l'hypothèse nulle à tort, i.e. de conserver H_0 alors que H_1 est vraie.

La première erreur est appelée erreur de première espèce. La deuxième erreur est appelée erreur de deuxième espèce.

1.4. Règle de décision et zone de rejet. La seule information dont dispose le statisticien pour aider à la décision, en l'espèce au diagnostic du patient, est l'expérience. Il lui faut donc bâtir une règle de décision permettant de déclarer, d'après la simple observation de l'expérience, s'il est préférable de conserver H_0 ou d'adopter H_1 . La règle de décision sera accompagnée d'un degré de fiabilité, spécifiant la confiance à lui accorder.

Dans le cas présent, nous allons raisonner par analogie avec les intervalles de confiance. Nous introduisons naturellement la moyenne empirique des observations, dont on sait qu'elle ressemble à la moyenne théorique. Nous posons de fait :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Il s'agit de fait de prendre une décision à partir de l'observation expérimentale \bar{x}_n de \bar{X}_n . Pour cela, nous introduisons une zone de rejet : si \bar{x}_n est trop éloignée de 1000, alors l'hypothèse H_0 est rejetée. Autrement dit, nous allons construire une zone de la forme $I =]-\infty, -c] \cup [c, +\infty[$ et lui associer la règle de décision suivante : si $\bar{x}_n - 1000 \in I$, H_0 est rejetée. Pour cette raison, I porte le nom de zone de rejet.

1.5. Construction de la zone de rejet. Pour construire la zone de rejet, il est nécessaire de savoir de combien la vraie moyenne et la moyenne empirique sont éloignées. En fait, nous savons que la distance entre \bar{X}_n et la vraie moyenne est liée à la variance. Précisément, nous savons que la loi de la statistique

$$T_n^\mu = n^{1/2} \frac{\bar{X}_n - \mu}{\sigma},$$

sous \mathbb{P}_μ ne dépend pas de μ et suit une loi $\mathcal{N}(0, 1)$. Il s'agit de la statistique que nous aurions utilisée pour construire un intervalle de confiance sur μ : il s'agit de la statistique que nous allons utiliser pour aider à la décision.

Autrement dit, pour un α donné, nous sommes capables de trouver un t^α tel que

$$\mathbb{P}_\mu\{-t^\alpha \leq T_n^\mu \leq t^\alpha\} = \alpha,$$

pour tout μ . Rappelons que t^α est le quantile d'ordre $2\alpha - 1$ de la gaussienne centrée réduite.

Supposons maintenant $\mu \in \Theta_0$, i.e. $\mu = \mu_0 = 1000$. Alors, nous sommes capables de dire que

$$\mathbb{P}_{1000}\left\{-t^\alpha \leq n^{1/2} \frac{\bar{X}_n - 1000}{\sigma} \leq t^\alpha\right\} = \alpha,$$

soit encore

$$\mathbb{P}_{1000}\{-t^\alpha \sigma n^{-1/2} \leq \bar{X}_n - 1000 \leq t^\alpha \sigma n^{-1/2}\} = \alpha.$$

De fait, nous choisissons $c = t^\alpha \sigma n^{-1/2}$, de sorte que la probabilité de déclarer H_1 à tort sous H_0 est exactement égale à α . Autrement dit, nous sommes capables de contrôler le risque de première espèce par α .

1.6. Dissymétrie des risques. Nous comprenons que la construction du test conduit à un contrôle de l'erreur de première espèce par un niveau de risque α préalablement fixé, appelé niveau du test. En pratique, α est choisi égal à 0,1 ou 0,05 voire 0,01. En pratique, le niveau de risque est fixé en fonction des conséquences que peut avoir la validation à tort de H_1 .

En revanche, la construction ne permet pas de contrôler, a priori, le risque de deuxième espèce. En général, on essaie de l'estimer a posteriori en contrôlant la puissance du test, i.e. la fonction

$$\mu \neq 1000 \mapsto \mathcal{P}(\mu) = \mathbb{P}_\mu\{\bar{X}_n - 1000 \in I\}.$$

Plus la puissance est proche de 1, plus le risque de déclarer H_0 à tort est petite, puisque, pour $\mu \neq 1000$,

$$\mathbb{P}_\mu\{\bar{X}_n - 1000 \notin I\} = 1 - \mathcal{P}(\mu).$$

Mathématiquement, le risque de deuxième espèce est défini par $\beta = \sup_{\mu \neq 1000} [1 - \mathcal{P}(\mu)]$.

2. Notions générales et autres exemples

2.1. Définition. Nous formalisons l'approche de la section précédente sous la forme :

Définition. Etant donné un modèle statistique $(\Omega, \mathbb{P}_\theta)_{\theta \in \Theta}$ muni d'une partition de Θ sous la forme $\Theta = \Theta_0 \cup \Theta_1$, une statistique T et un intervalle I , la statistique

$$\Delta = \mathbf{1}_I(T)$$

un test de l'hypothèse nulle $H_0 = \theta \in \Theta_0''$ contre l'hypothèse alternative $H_1 = \theta \in \Theta_1''$ de niveau inférieur à $\alpha \in]0, 1[$ et de zone de rejet I si

$$\forall \theta \in \Theta_0, \quad \mathbb{P}_\theta\{T \in I\} \leq \alpha.$$

La puissance du test est donnée par la fonction :

$$\mathcal{P} : \theta \in \Theta_1 \mapsto \mathbb{P}_\theta\{T \in I\}.$$

2.2. Mode opératoire. La construction d'un test s'effectue en plusieurs étapes :

- (1) Choix du modèle
- (2) Identification des hypothèses
- (3) Choix du niveau du test
- (4) Choix de la statistique utilisée pour la décision
- (5) Choix de la forme de la zone de rejet
- (6) Calcul explicite de la zone de rejet
- (7) Décision finale
- (8) Calcul éventuel de la puissance

2.3. Un deuxième exemple. Considérons un deuxième exemple pour valider le mode opératoire. Des crânes ont été découverts dans une grotte et leurs largeurs en mm sont :

141	140	145	135	147	141	145
-----	-----	-----	-----	-----	-----	-----

De nombreux crânes précédemment découverts dans la région de cette grotte ont une largeur moyenne égale à 146. En supposant une loi normale pour la largeur des crânes, décider au niveau 0.1 si les crânes découverts proviennent ou non de la même population.

Mettons en place le test.

Choix du modèle. On utilise le modèle gaussien $(\Omega, \mathcal{N}(m, \sigma^2)^{\otimes n})_{m \in \mathbb{R}, \sigma > 0}$ avec $n = 7$. Ici m et σ sont en mm. Le modèle est muni des statistiques d'observation X_1, \dots, X_n .

Choix des hypothèses $H_0 = "m = m_0 = 146"$ contre $H_1 = "m \neq m_0"$.

Niveau du test $\alpha = .1$.

Statistique utilisée La décision se fait à partir de l'observation de la moyenne empirique \bar{X}_n et plus précisément de sa distance à m_0 .

Attention : cette distance est liée à la variance ici inconnue ! Nous comprenons de fait que la règle d'observation doit faire intervenir la variance empirique. Autrement dit, nous allons fonder la décision sur l'observation de la réalisation de la statistique

$$T = \sqrt{n} \frac{\bar{X}_n - 1000}{\bar{\Sigma}_n}$$

Forme de la zone de rejet $I = (-\infty, -c] \cup [c, \infty)$. Le risque de première espèce est :

$$\alpha = \mathbb{P}_{m_0} \{|T| \geq c\}.$$

Calcul de la zone de rejet Sous H_0 , on sait que T suit une loi de Student de niveau $n - 1 = 6$. De fait, il s'agit de chercher c tel que la probabilité que la loi de Student de degré 6 soit dans $] -c, c[$ soit égale à 0.9. Par utilisation de la table, nous obtenons $c = 1,943$.

Décision finale Il faut maintenant calculer la réalisation de la statistique utilisée pour le test, i.e. $\sqrt{6}(\bar{x}_n - 146)/\bar{\sigma}_n$. Nous trouvons -2.69 . Le résultat est dans la zone de rejet : nous acceptons l'hypothèse H_1 avec un niveau de risque égal à 10%.

3. Tests unilatères

Dans les exemples précédents, nous avons testé l'égalité d'un paramètre avec une valeur donnée contre la différence du paramètre avec cette même valeur.

Il existe des situations où le test vise à vérifier (éventuellement) une borne sur un paramètre. Nous parlerons de tests unilatères.

3.1. Test unilatère sur la moyenne. Donnons un exemple : un organisme a demandé au ministère d'agréer une source d'eau afin de la distribuer dans le commerce. En un point de captage, ont été répétées six mesures du taux d'oxygène dans l'eau (en parties par million). On a trouvé :

4.92	5.10	4.93	5.02	5.06	4.71
------	------	------	------	------	------

La norme en dessous de laquelle la consommation est interdite est de 5 ppm. Au vu des observations, est-il possible de déclarer l'eau potable ?

Il s'agit ici de mettre en place un modèle gaussien constitué de $n = 6$ mesures. Le test porte alors sur $H_0 : \mu \leq \mu_0 = 5$ contre $H_1 : \mu > \mu_0 = 5$. En effet, il est essentiel de contrôler la probabilité de déclarer le source potable à tort.

Le niveau du test peut être fixé à 0.05 ou 0.01.

La statistique utilisée est la même que dans le cadre d'un test bilatéral. Autrement dit, tout se passe encore autour de la valeur μ_0 .

La zone de rejet est de la forme $I = [c, +\infty)$. Le risque de première espèce est donnée par

$$\alpha = \sup_{\mu \leq \mu_0} \mathbb{P}_\mu \{T \geq c\}.$$

Le problème est que nous ne connaissons pas la loi de T sous \mathbb{P}_μ , pour $\mu \leq \mu_0$. Mais nous pouvons écrire

$$\begin{aligned} T &= \sqrt{n} \frac{\bar{X}_n - \mu}{\bar{\Sigma}_n} + \sqrt{n} \frac{\mu - \mu_0}{\bar{\Sigma}_n} \\ &\leq \sqrt{n} \frac{\bar{X}_n - \mu}{\bar{\Sigma}_n} \end{aligned}$$

De fait, nous comprenons que

$$T \geq c \Rightarrow \sqrt{n} \frac{\bar{X}_n - \mu}{\bar{\Sigma}_n} \geq c.$$

Nous en déduisons

$$\alpha \leq \sup_{\mu \leq \mu_0} \mathbb{P}_\mu \left\{ \sqrt{n} \frac{\bar{X}_n - \mu}{\bar{\Sigma}_n} \geq c \right\}.$$

Mais, toutes ces quantités sont égales ! Comme l'égalité est vraie en $\mu = \mu_0$, nous choisissons c comme le $1 - \alpha$ quantile de la loi de Student de degré 5.

A.N. : En choisissant $\alpha = 0.05$, nous obtenons $c = 2.015$. Nous pouvons calculer $\bar{x}_n = 4.96$ et $\bar{\sigma}_n = 0.14$ et $t = -0.75$. L'hypothèse H_0 est conservée.

3.2. Test unilatère sur l'écart-type. On veut tester la précision d'une méthode de mesure d'alcoolémie sur un échantillon sanuin. La précision est définie comme étant deux fois l'écart-type de la loi.

On partage un échantillon de référence en 6 éprouvettes que l'on soumet à analyse d'un laboratoire. Les valeurs trouvées sont en g par litres

1.35	1.26	1.48	1.32	1.50	1.44
------	------	------	------	------	------

Tester l'hypothèse nulle : « la précision est inférieure à 0.1 gramme par Litre au niveau 0.05 ».

Ici, le modèle est un modèle gaussien de la forme $(\Omega, \mathcal{N}^{\otimes n}(m, \sigma^2))_{m \in \mathbb{R}, \sigma > 0}$ à moyenne et variance inconnues avec $n = 6$. L'hypothèse H_0 s'écrit « $\sigma \leq 0.05$ » et l'hypothèse H_1 s'écrit « $\sigma > 0.05$ ». Le niveau du test est $\alpha = 0.05$. Nous allons fonder le test sur la statistique

$$T = (n - 1) \frac{\bar{\Sigma}_n^2}{\sigma_0^2}.$$

La zone de rejet est de la forme $[0, c]$: nous rejettons pour des valeurs élevées de la variance empirique et donc des petites valeurs de son inverse.

Pour calculer c nous remarquons que la loi de T sous $\mathbb{P}_{m,\sigma_0^2}$ est exactement la loi du χ^2 à $(n-1)$ degrés de liberté. Si $\sigma < \sigma_0$, nous remarquons que

$$T \geq (n-1) \frac{\bar{\Sigma}_n^2}{\sigma^2},$$

de sorte que

$$\mathbb{P}_{m,\sigma^2}\{T \leq c\} \leq \mathbb{P}_{m,\sigma_0^2}\left\{(n-1) \frac{\bar{\Sigma}_n^2}{\sigma_0^2} \leq c\right\},$$

avec égalité si $\sigma = \sigma_0$. Toutes les quantités dans le terme de droite ci-dessus sont égales : il s'agit de la valeur au point c de la fonction de répartition de la loi du χ^2 à $(n-1)$ degrés de liberté. Il suffit donc de poser c égal au 0.05 quantile de la loi du χ^2 à $n-1$ degrés de liberté. Nous obtenons $c = 1.15$.

A.N. Nous calculons $\bar{\sigma}_n = .096$ et $t = 18,4$. Nous rejettons l'hypothèse H_0 au niveau de risque 0.05.

3.3. Test unilatère non gaussien. Un centre d'appel téléphonique prétend que le temps moyen d'attente à son service est inférieur à trois minutes.

Voici les temps d'attente constatés par une association de consommateurs (rangés par ordre croissant).

0.75	1.25	2.25	1.75	4	6
------	------	------	------	---	---

Tester, au niveau 0,05, l'hypothèse H_0 : le temps moyen d'attente est inférieur à deux minutes contre l'hypothèse alternative au niveau 0.05. (L'association de consommateurs veut être sûre de son verdict en cas de mise en défaut du centre d'appels.)

Utiliser la statistique

$$\lambda_0 n \min(X_1, \dots, X_n),$$

avec $\lambda_0 = 1/2$, $n = 6$ et X_1, \dots, X_n statistiques d'observation.

Le modèle statistique utilisé est le modèle exponentiel $(\Omega, \text{Exp}^{\otimes n}(\lambda))_{\lambda > 0}$, avec $n = 6$. L'hypothèse H_0 s'écrit $\lambda^{-1} \leq 2$. Le niveau du test est choisi égal à $\alpha = 0,05$.

On choisit la statistique de test proposé. La zone de rejet est de la forme $[c, +\infty)$: on rejette pour des grandes valeurs des observations.

Sous \mathbb{P}_{λ_0} , T suit une loi exponentielle de paramètre 1. Il suffit en effet de calculer la fonction de survie : $\mathbb{P}_{\lambda_0}\{n\lambda_0 \min(X_1, \dots, X_n) > t\} = (\exp(-t/n))^n = \exp(-t)$.

Maintenant, si $\lambda \geq \lambda_0$, alors

$$T \leq \lambda n \min(X_1, \dots, X_n),$$

et nous voyons que $\mathbb{P}_{\lambda}\{T \geq c\}$ est inférieur à $\mathbb{P}_{\lambda}\{\lambda n \min(X_1, \dots, X_n) \leq c\} = \exp(-c)$. Nous choisissons donc $\exp(-c) = \alpha$.

A.N. $c = -\ln(0.1) = 2.30$ et $t = 1.5$. On conserve H_0 .