Introduction à la statistique inférentielle

1 Introduction

1.1 Quelques généralités

La démarche statistique consiste à traiter et à interpréter les informations recueillies par le biais de données. Elle comporte quatre grands aspects : le recueil des données, l'aspect descriptif ou exploratoire, l'aspect inférentiel ou décisionnel et la modélisation statistique.

Le recueil des données Cette étape est importante car elle doit permettre d'obtenir des données de "bonne qualité" en un certain sens. Contrairement à ce qu'indique le vocabulaire, les informations dont a besoin le statisticien ne sont pourtant pas "données" et la qualité des résultats obtenus dépendra autant de la manière dont les données ont été collectées que la méthode statistique utilisée. La théorie des sondages et celle des plans d'expériences fournissent un cadre théorique pour la recherche de données optimales.

La statistique exploratoire ou descriptive Son but est de synthétiser et de résumer l'information contenue dans les données. Elle utilise pour cela des représentations des données sous forme de tableaux, de graphiques ou d'indicateurs numériques (tels que la moyenne, la variance, la corrélation linéaire, ...pour des variables quantitatives). Cette phase est connue sous le nom de **statistique descriptive**. On parle de statistique descriptive *univariée* lorsque l'on regarde une seule variable, de statistique descriptive *bivariée* lorsque l'on regarde simultanément deux variables, et de statistique descriptive *multidimensionnelle* lorsque l'on regarde simultanément p variables. Dans ce dernier cas, on parle aussi d'**analyse des données**.

La statistique inférentielle Son but est d'étendre (d'inférer) les propriétés constatées sur l'échantillon (grâce l'analyse exploratoire par exemple) à la population toute entière, et de valider ou d'infirmer des hypothèses. Contrairement à la statistique exploratoire, des hypothèses probabilistes sont ici nécessaires : elle suppose un modèle probabiliste. L'estimation ponctuelle ou par intervalle de confiance et la théorie des tests d'hypothèses constituent une partie principale de la statistique inférentielle.

La modélisation statistique Elle consiste en général à rechercher une relation "approximative" entre une variable et plusieurs autres variables, la forme de cette relation est le plus souvent linéaire. Lorsque la variable à expliquer est quantitative et que les variables explicatives sont aussi quantitatives, on parle de *régression linéaire*. Si les variables explicatives sont qualitatives, on parle alors d'*analyse de la variance*. Le *modèle linéaire général* englobe une grande partie de tous les cas de figures possibles.

Remarque D'un point de vue méthodologique, on notera que la statistique descriptive précède en général la statistique inférentielle et la modélisation statistique dans une démarche de traitement de données : ces différents aspects de la statistique se complètent bien plus qu'ils ne s'opposent.

1.2 Modèle d'échantillonage

Soit $X = (X_1, \ldots, X_n)$, un vecteur de n variables aléatoires réelles X_i que l'on supposera *indépendantes* et *identiquement distribuées* (i.i.d.). Soit P la loi de X_1.

Hypothèse fondamentale On va supposer que la loi de probabilité P appartient à une famille de lois de probabilité $\mathcal{P}$.

Définition. On appelle *modèle statistique* (ou structure statistique) le triplet $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathcal{P})^n$, c'est à dire la donnée d'une famille de lois de probabilité $\mathcal{P}$ à laquelle on impose que P appartient.

Définition. On appelle *échantillon* une réalisation $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$ d'un modèle statistique.

Remarque La connaissance du phénomène étudié permet d'avoir une idée pour le choix de la famille de lois de probabilité $\mathcal{P}$: cette connaissance peut provenir de l'étude de la série statistique $(x_1, \ldots, x_n)$, de l'expérience du statisticien, ...

Le modèle d'échantillonnage paramétrique Il consiste, dans le cadre de l'échantillonnage, à supposer que la famille $\mathcal{P}$ de lois de probabilité est indicée par un paramètre θ. On note alors

$$\mathcal{P} = (P_\theta,\ \theta \in \Theta \subset \mathbb{R}^p),$$

avec :

- P_θ est la loi de probabilité correspondant à la valeur θ du paramètre.
- Θ est l'espace paramétrique (dans lequel θ peut prendre sa valeur).
- p est la dimension du paramètre (pour $p = 1$, on parle de paramètre unidimensionnel, pour $p > 1$, on parle de paramètre multidimensionnel ou vectoriel).

Notation. Un modèle d'échantillonnage paramétrique sera donc noté

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}^p))^n .$$

En dehors de ce cadre, on parle de modèle non paramétrique.

1.3 Définition d'une statistique

C'est une notion fondamentale pour la suite. Reprenons le modèle statistique général $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathcal{P})^n$.

On appelle *statistique* toute variable aléatoire définie sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ à valeurs dans $(\mathbb{R}^q, \mathcal{B}_{\mathbb{R}^q})$. Très souvent, les statistiques intervenant dans le modèle paramétrique auront pour objet de "préciser" le paramètre $\theta \in \mathbb{R}^p$. Le plus souvent, on aura $q = p$.

Une statistique est une variable aléatoire que l'on notera T_n. Elle ne doit pas dépendre de la famille de lois de probabilité $\mathcal{P}$. En particulier, dans un modèle paramétrique, une statistique T_n ne doit pas dépendre de θ. Mais la loi de probabilité de T_n va dépendre de θ.

Cas usuel : le modèle d'échantillonnage paramétrique réel.

$$\begin{array}{rcccc} T_n & : & (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}^p))^n & \longrightarrow & (\mathbb{R}^q, \mathcal{B}_{\mathbb{R}^q}) \\ & & (X_1, \ldots, X_n) & \longmapsto & T_n(X_1, \ldots, X_n) \end{array}$$

Exemple Prenons $P_\theta = N(\mu, 1)$. On a ici $\theta = \mu$ (avec $p = 1$). Une statistique permettant de "préciser" le paramètre θ (moyenne d'une loi normale réduite) peut être naturellement la variable aléatoire

$$T_n(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n,$$

moyenne empirique des X_i (ici $q = 1$).

Remarque Il convient bien de distinguer :

- la statistique T_n qui est une variable aléatoire définie sur $(\mathbb{R}, \mathcal{B}_\mathbb{R}, \mathcal{P})^n$. Dans l'exemple précédent, c'est l'"opérateur" moyenne empirique (on est ici au niveau conceptuel, mathématique).
- la variable aléatoire $T_n(X_1, \ldots, X_n)$ qui est une variable aléatoire sur $(\mathbb{R}^q, \mathcal{B}_{\mathbb{R}^q})$. Dans l'exemple précédent, c'est la variable aléatoire $\overline{X}_n$ (on est ici au niveau de la statistique inférentielle).
- la valeur observée de cette variable aléatoire : $T_n(x_1, \ldots, x_n) \in \mathbb{R}^q$. Dans l'exemple précédent, c'est le nombre réel $\mathcal{B}arx_n$, moyenne empirique de la série statistique $x_1, \ldots, x_n$ (on est au niveau de la statistique descriptive).

1.4 Notions d'estimation

Prenons le modèle le plus usuel, à savoir le modèle d'échantillonnage paramétrique réel

$$(\mathbb{R}, \mathcal{B}_\mathbb{R}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}^p))^n .$$

Si le paramètre θ est connu, alors la loi de probabilité P_θ est connue. Les méthodes standards de la statistique inférentielle (estimation et test d'hypothèse) ont pour objectif de préciser au maximum le paramètre θ quand ce dernier est inconnu.

Estimation Ce sont de méthodes permettant de fixer une valeur (on parle d'*estimation ponctuelle*) ou un ensemble de valeurs (on parle d'*estimation par intervalle de confiance*) pour le paramètre θ.

2 Estimation ponctuelle d'un paramètre unidimensionnel

On considère dans cette partie un modèle paramétrique réel d'échantillonnage :

$$(\mathbb{R}, \mathcal{B}_\mathbb{R}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}))^n .$$

2.1 Définition de la vraisemblance

Si la loi P_θ est une loi de probabilité continue, elle se caractérise par une densité de probabilité : $f_\theta\ :\ \mathbb{R}to\mathbb{R}^+$ vérifiant $\int_\mathbb{R} f_\theta(x)dx = 1$. Si la loi P_θ est discrète et si l'ensemble des valeurs possibles de X_1 est un sous-ensemble D de $\mathbb{R}$ fini ou dénombrable, la loi P_θ est caractérisée par les quantités $P_\theta[X_1 = x]$ pour $x \in D$, c'est à dire que $P_\theta\ :\ D \subset \mathbb{R} \to [0, 1]$ vérifiant $\sum_{x \in D} P_\theta[X_1 = x] = 1$.

Définition. Pour faire jouer au paramètre θ le rôle d'une variable (au sens mathématique), on définit une fonction très générale appelée *vraisemblance* et notée L :

$$\begin{array}{rccl} L & : & \mathbb{R}^n \times \Theta & \longrightarrow & \mathbb{R}^+ \\ & & (x_1, \ldots, x_n, \theta) & \longmapsto & L(x_1, \ldots, x_n, \theta) = \left\{ \begin{array}{ll} \prod_{i=1}^n f_\theta(x_i) & \text{si } P_\theta \text{ continue} \\ \prod_{i=1}^n P_\theta[X_i = x_i] & \text{si } P_\theta \text{ discrète} \end{array} \right. \end{array}$$

Quelques commentaires.

— La vraisemblance sera systématiquement notée L (comme *likelihood* en anglais).
— On écrira $L(x_1, \ldots, x_n, \theta) = L(x, \theta) = L$ quand il n'y aura pas d'ambiguïté.
— La vraisemblance spécifie (caractérise) le modèle et ne peut donc être précisée que si le modèle a été choisi.
— Une vraisemblance vérifie toujours la contrainte suivante :

$$\int_{\mathbb{R}^n} L(x, \theta) dx = 1, \ \forall \theta \in \Theta.$$

(Par convention, dans le cas discret, on remplace $\int_{\mathbb{R}^n}$ par $\sum_{x \in D}$.)
— Par la suite, on utilisera souvent la notation $L(X_1, \ldots, X_n, \theta) = L(X, \theta)$ qui est une variable aléatoire réelle à valeurs dans $\mathbb{R}^+$.
— $L(X_1, \ldots, X_n, \theta)$ n'est pas une statistique car elle dépend de θ. Cependant, elle permettra de contruire des statistiques (par exemple, voir la méthode du maximum de vraisemblance).
— On utilisera souvent la *log-vraisemblance* définie comme le logarithme népérien de la vraisemblance : $\ln\ L$.

2.2 Estimateurs

Définition On appelle *estimateur du paramètre* θ toute statistique à valeur dans $(\Theta, \mathcal{B}_\Theta)$, il sera généralement noté T_n. On appelle *estimation* la valeur observée de T_n sur l'échantillon, elle sera notée $T_n(x_1, \ldots, x_n) \in \Theta$.

Définition On appelle *biais* de l'estimateur T_n pour le paramètre θ la quantité

$$B(T_n) = \mathbb{E}[T_n] - \theta.$$

On appelle *estimateur sans biais de* θ un estimateur T_n tel que $B(T_n) = 0$, sinon on parle d'estimateur biaisé. Si l'estimateur T_n est biaisé, mais que $B(T_n) \to 0$ quand $n \to +\infty$, on dit que T_n est *asymptotiquement sans biais* pour θ.

Définition On dit que T_n est *convergent* pour θ s'il converge en probabilité vers θ :

$$\forall \epsilon > 0, \ P(|T_n - \theta| < \epsilon) \longrightarrow 1 \text{ quand } n \to +\infty.$$

Proposition

(i) Si T_n est un estimateur sans biais de θ et si $\mathbb{V}(T_n) \longrightarrow 0$ quand $n \to +\infty$, alors T_n est un estimateur convergent pour θ.
(ii) Si T_n est un estimateur asymptotiquement sans biais de θ et si $\mathbb{V}(T_n) \longrightarrow 0$ quand $n \to +\infty$, alors T_n est un estimateur convergent pour θ.

2.3 Comparaisons des estimateurs

On utilise le *risque quadratique* pour comparer deux estimateurs du même paramètre θ. Pour l'estimateur T_n de θ, il est defini par :

$$R(T_n, \theta) = \mathbb{E}[(T_n - \theta)^2].$$

Propriété. $R(T_n, \theta) = (B(T_n))^2 + \mathbb{V}(T_n)$.

Commentaires.
- L'idée est de minimiser le risque quadratique $R(T_n, \theta)$, c'est à dire mimiser le biais (si possible l'annuler) ainsi que la variance de l'estimateur.
- La variance de T_n ne pourra pas descendre en dessous d'une certaine borne (voir l'inégalité de Cramer-Rao).
- La quantité $R(T_n, \theta)$ est aussi appelé *erreur quadratique moyenne*.
- Le risque quadratique est le critère de choix généralement utilisé entre deux estimateurs.

2.4 Propriétés des statistiques $\overline{X}_n$, V_n^2 et S_n^2

Soient n variables aléatoires réeelles $X_1, \ldots, X_n$ indépendantes et de même loi (quelconque), admettant une espérance μ, une variance σ^2, ainsi que des moments centrés d'ordre 3 et 4 (notés μ_3 et μ_4).

On définit les estimateurs

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \quad \text{et} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Propriétés générales.
- $\mathbb{E}[\overline{X}_n] = \mu$; $\mathbb{V}(\overline{X}_n) = \dfrac{\sigma^2}{n}$.
- $\mathbb{E}[S_n^2] = \sigma^2$; $\mathbb{V}(S_n^2) = \dfrac{\mu_4}{n} - \dfrac{n-3}{n(n-1)}\sigma^4$,
  avec $\mu_4 = \mathbb{E}[(X_i - \mu)^4]$ et $\sigma^4 = (\sigma^2)^2$.
- $\mathbb{E}[V_n^2] = \dfrac{n-1}{n}\sigma^2$; $\mathbb{V}(V_n^2) = \dfrac{(n-1)^2}{n^3}\mu_4 - \dfrac{(n-1)(n-3)}{n^2}\sigma^4$.
- $\mathrm{cov}(\overline{X}_n, S_n^2) = \dfrac{\mu_3}{n}$,
  avec $\mu_3 = \mathbb{E}[(X_i - \mu)^3]$.
- $\mathrm{cov}(\overline{X}_n, V_n^2) = \dfrac{n-1}{n^2}\mu_3$.

Cas particulier des lois normales. Si l'on suppose en plus que chaque variable aléatoire réelle X_i suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, alors on a :
- $\mathbb{V}(S_n^2) = \dfrac{2}{n-1}\sigma^4$.
- $\mathbb{V}(V_n^2) = \dfrac{2(n-1)}{n^2}\sigma^4$.
- $\mathrm{cov}(\overline{X}_n, S_n^2) = \mathrm{cov}(\overline{X}_n, V_n^2) = 0$.

On peut de plus démontrer les résultats fondamentaux suivants :

— $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
— $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$.
— Les statistiques $\overline{X}_n$ et S_n^2 sont indépendantes.

3 Inégalité de Cramer-Rao

On reste dans le cadre d'un modèle paramétrique réel d'échantillonnage

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}^p))^n .$$

Soit $\mathcal{X} \subset \mathbb{R}$ le support de $f(x,\theta)$ où

$$\mathcal{X} = \{x \in \mathbb{R} \mid f(x,\theta) > 0\}.$$

Notons que $\mathcal{X}$ peut dépendre ou non de θ. Le support de la vraisemblance L est alors $\mathcal{X}^n$.

On est amené à faire un certain nombre d'*hypothèses de régularité* sur f pour pouvoir établir certaines propriétés.

(H1) $\mathcal{X}$ est indépendant de θ.

(H2) Θ est un ouvert.
Le fait d'avoir aussi $f(x,\theta) > 0, \forall (x,\theta) \in \mathcal{X} \times \Theta$, va permettre de ne pas avoir de problèmes de différentiabilité aux frontières.

(H3) $\frac{\partial}{\partial\theta} f(x,\theta)$ et $\frac{\partial^2}{\partial\theta^2} f(x,\theta)$ sont définies $\forall (x,\theta) \in \mathcal{X} \times \Theta$.

(H4) On peut dériver deux fois $f(x,\theta)$ selon θ sous le signe $\int_{\mathbb{R}} dx$,
ce qui impose que l'on a par exemple $\frac{\partial}{\partial\theta} \int_{\mathbb{R}} f(x,\theta)dx = \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f(x,\theta)dx$.

Définition. On appelle *fonction score* ou *score* la fonction S définie par :

$$\begin{array}{rccl} S & : & \mathcal{X} \times \Theta & \longrightarrow & \mathbb{R} \\ & & (x,\theta) & \longmapsto & S(x,\theta) = \frac{\partial}{\partial\theta} \ln(f(x,\theta)) \end{array}$$

Commentaires.
— Le score n'est défini que si **(H1)**, **(H2)** et **(H3)** sont vraies.
— Le score intervient très souvent en estimation.
— On peut définir de même le score à partir de la vraisemblance. On le notera alors :

$$S_n(x,\theta) = \frac{\partial}{\partial\theta} \ln(L(x,\theta)) \ \text{ avec ici } x \in \mathbb{R}^n.$$

Dans le modèle d'échantillonnage, on a : $S_n(x,\theta) = \sum_{i=1}^n S(x_i,\theta)$.

Propriété. Supposons **(H4)** vraie, on a alors :

$$\mathbb{E}[S(X,\theta)] = 0 \quad \text{et} \quad \mathbb{E}[S_n(X,\theta)] = 0.$$

Définition. On appelle *information de Fisher* la fonction I définie par :

$$\begin{array}{rccl} I & : & \Theta & \longrightarrow & \mathbb{R}^+ \\ & & \theta & \longmapsto & I(\theta) = \mathbb{E}\left[(S(X,\theta))^2\right] \end{array}$$

Commentaires.

— L'information de Fisher I n'est définie que si les hypothèses **(H1)**, **(H2)** et **(H3)** sont vérifiées.

— L'information de Fisher I ne dépend pas de X, c'est à dire de l'échantillon.
Elle ne dépend que de θ et du modèle choisi. C'est une information contenue dans le modèle sur le paramètre θ.

— On peut aussi poser (en terme de vraisemblance) :

$$I_n(\theta) = \mathbb{E}\left[(S_n(X,\theta))^2\right].$$

Propriété. Si **(H4)** est vraie, alors

$$I(\theta) = \mathbb{V}(S(X,\theta)) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln(f(X,\theta))\right] \text{ et } I_n(\theta) = \mathbb{V}(S_n(X,\theta)) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln(L(X,\theta))\right].$$

Dans le modèle d'échantillonnage, on a donc : $I_n(\theta) = nI(\theta)$.

On a maintenant besoin d'une hypothèse supplémentaire :
(H5) $0 < I_n(\theta) < +\infty$ ou $0 < I(\theta) < +\infty$.

Soit T_n un estimateur quelconque de θ. Posons $\mathbb{E}[T_n] = g(\theta)$.

Inégalité de Cramer-Rao. Si les hypothèses **(H1)**, **(H2)**, **(H3)**, **(H4)** et **(H5)** sont vérifiées, alors on a :

$$\mathbb{V}(T_n) \geq \frac{\left[\frac{\partial}{\partial\theta}g(\theta)\right]^2}{I_n(\theta)}.$$

Commentaires.

— La partie de droite est appelée *borne inférieure de l'inégalité de Cramer-Rao*. Nous la noterons $K_{T_n}(\theta)$.

— L'estimateur T_n intervient au numérateur.
Au dénominateur, seul le modèle intervient.

— Dans le cas particulier des estimateurs sans biais (soit $\mathbb{E}(T_n) = \theta$), on a :

$$\mathbb{V}(T_n) \geq I_n^{-1}(\theta).$$

Définition

— On dit que l'estimateur T_n est *efficace* s'il vérifie $\mathbb{V}(T_n) = K_{T_n}(\theta)$.
Notons que cette notion dépend de l'estimateur et du modèle.

— Si T_n n'est pas efficace mais que $\dfrac{K_{T_n}(\theta)}{\mathbb{V}(T_n)} \to 1$ quand $n \to +\infty$, on dit que l'estimateur T_n est *asymptotiquement efficace*.

Propriétés.

(P1) Si T_n est un estimateur efficace de θ, alors $kT_n + b$ est aussi un estimateur efficace de θ, $\forall k \in \mathbb{R}^*, \forall b \in \mathbb{R}$.

(P2) Soient T_{1n} et T_{2n} deux estimateurs sans biais du paramètre θ. S'ils sont tous les deux efficaces, alors $T_{1n} = T_{2n}$.

Remarques.

— Si T_n est un estimateur efficace de θ, il apparaît alors intéressant de chercher dans les estimateurs de la forme kT_n celui qui minimise (ou annule) le biais. Par contre, on prend toujours $b = 0$: en effet, la constante b ne dépend pas de X (donc de l'expérience), ni de θ (sinon cela ne serait plus un estimateur).

— Pour l'instant, on ne peut rien dire de deux estimateurs efficaces d'un même paramètre qui n'auraient pas le même biais.

Supposons que, dans le modèle "usuel" (c'est à dire le modèle d'échantillonnage paramétrique réel, avec un paramètre unidimensionnel, vérifiant les hypotèses **(H1)** à **(H5)**), on ait un estimateur T_n efficace de θ.

Définition. On appelle *famille exponentielle à paramètre unidimensionnel* θ toute loi de probabilité (discrète ou continue) dont la densité (ou la probabilité) peut se mettre sous la forme :

$$f(x,\theta) = \begin{cases} \exp[\alpha(\theta)\beta(x) + \gamma(\theta) + \delta(x)] & \text{si} \quad x \in \mathcal{X} \\ 0 & \text{si} \quad x \notin \mathcal{X} \end{cases}$$

Théorème. Un estimateur T_n d'un paramètre θ est efficace si et seulement si

(i) le modèle appartient à la famille exponentielle ;

(ii) la statistique T_n s'écrit sous la forme $T_n = k\sum_{i=1}^{n} \beta(X_i)$.

Remarques.

— Dans la définition de la famille exponentielle, la définition des fonctions $\alpha(.)$, $\beta(.)$, $\gamma(.)$ et $\delta(.)$ n'est pas unique. En particulier, on peut poser $\alpha'(.) = \frac{1}{a}\alpha(.)$, $\beta'(.) = a\beta(.)$, $\gamma'(.) = \gamma(.) - b$ et $\delta'(.) = \delta(.) + b$ pour $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$, et on obtiendra la même densité.

— Un estimateur de la forme donnée dans le théorème n'a pas forcément d'utilité concrète.

4 Notion d'exhaustivité

On reste dans le cadre d'un modèle paramétrique réel d'échantillonnage

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}^p))^n .$$

De manière intuitive, une statistique T_n sera exhaustive pour le paramètre θ si elle résume toute l'information sur la paramètre contenue dans l'échantillon.

On se restreint ici aux estimateurs. La statistique T_n est ici une variable aléatoire définie sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, (P_\theta,\ \theta \in \Theta \subset \mathbb{R}))^n$ à valeurs dans $(\Theta, \mathcal{B}_\Theta)$. On considère :

— la densité (ou la probabilité) de T_n que l'on va noter $\psi_n(t,\theta)$,

— la densité conjointe de l'échantillon $(X_1, \ldots, X_n)$ notée $L(x_1, \ldots, x_n, \theta)$,

— la densité conjointe de $(X_1, \ldots, X_n)$ et de T_n que l'on va noter $Ł(x_1, \ldots, x_n, t, \theta)$.
On peut remarquer qu'il y a clairement dans Ł une certaine redondance entre les x_i et t, mais on conservera cette notation par souci de clarté.
En introduisant Ψ la densité conditionnelle des X_i sachant que $T_n = t$, on peut écrire

$$Ł(x_1, \ldots, x_n, t, \theta) = \psi_n(t,\theta)\Psi(x_1, \ldots, x_n, \theta | T_n = t).$$

Définition On dit que T_n *est exhaustive pour* θ si et seulement si Ψ ne dépend pas de θ.

En général, la définition de l'exhaustivité n'est pas très commode pour vérifier l'exhaustivité d'une statistique. Le théorème de factorisation (admis) sera souvent plus pratique.

Théorème. La statistique T_n est exhaustive pour θ si et seulement s'il existe deux applications $h(\cdot,\cdot)$ et $k(\cdot)$ telle que :

$$\begin{aligned} h &: \Theta \times \Theta \longrightarrow \mathbb{R}^+ \\ k &: \mathbb{R}^n \longrightarrow \mathbb{R}^+ \end{aligned}$$

et la vraisemblance $L(x_1,\ldots,x_n,\theta) = h[T_n(x_1,\ldots,x_n,\theta),\theta]k(x_1,\ldots,x_n)$.

Application à la famille exponentielle. Montrer que, dans la famille exponentielle, toute statistique de la forme $T_n = k\sum_{i=1}^n \beta(X_i)$ est exhaustive pour θ.

5 Quelques méthodes usuelles d'estimation

On va présenter brièvement ici des méthodes permettant de calculer de manière systématique un estimateur pour le paramètre θ d'un modèle statistique.

5.1 Méthode empirique

Si le paramètre θ considéré représente une quantité particulière pour le modèle (par exemple, l'espérance ou la variance), on peut naturellement choisir comme estimateur la quantité empirique correspondante pour l'échantillon $X_1,\ldots,X_n$.

Exemple. Lorsque $\theta = \mathbb{E}[X]$, on peut choisir $\overline{X}_n$ comme estimateur de θ. Lorsque $\theta = \mathbb{V}(X)$, on peut choisir V_n^2 comme estimateur de θ ; on peut ensuite modifier l'estimateur pour améliorer ses qualités (V_n^2 asymptotiquement sans biais $\longrightarrow$ S_n^2 sans biais).

Remarque. Cette méthode très naturelle est relativement limitée.

5.2 Méthode des moindres carrés

Elle est utilisable lorsque l'espérance de la loi est une fonction inversible du paramètre θ, c'est à dire $\mathbb{E}[X] = h(\theta)$ où $h(.)$ est bijective.

Définition. On appelle *estimateur des moindres carrés de* θ la statistique

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{i=1}^n (X_i - h(\theta))^2.$$

Propriété. $\hat{\theta}_n = h^{-1}\left(\overline{X}_n\right)$.

Exemple. Lorsque $\theta = \mathbb{E}[X]$, on a alors $h = Id = h^{-1}$. On en déduit que l'estimateur des moindres carrés $\hat{\theta}_n$ est : $\hat{\theta}_n = \overline{X}_n$.

Remarques.

— Cette méthode s'adapte au cas d'un paramètre θ multidimensionnel.

— Cette méthode est essentiellement utilisée dans le cadre du modèle linéaire. Cela donne les mêmes résultats que la méthode du maximum de vraisemblance.

5.3 Méthode des moments

Rappels.

— Soit X une variable aléatoire réelle.
On appelle *moment (théorique) d'ordre r* : $M_r = \mathbb{E}[X^r]$.
On appelle *moment (théorique) centré d'ordre r* : $\overline{M}_r = \mathbb{E}[(X - \mathbb{E}[X])^r]$.

— Soit $(x_1, \ldots, x_n)$ les valeurs observées d'un échantillon de taille n.
On appelle *moment empirique d'ordre r* : $m_r = \frac{1}{n}\sum_{i=1}^n (x_i)^r$.
On appelle *moment empirique centré d'ordre r* : $\overline{m}_r = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^r$.

Principe. Supposons le paramètre θ de dimension p. La méthode consiste à poser un système d'équations en égalant moments théoriques (centrés ou non) et moments empiriques :

$$\left\{ \begin{array}{rcl} M_1(\theta) & = & m_1 \\ & \vdots & \\ M_p(\theta) & = & m_p \end{array} \right.$$

Chaque $M_j(\theta)$ dépend de θ, alors que les m_j n'en dépendent pas. On a donc un système à p équations et p inconnues pour obtenir les estimateurs des p composantes de θ.

Exemple. Soit une variable aléatoire réelle admettant une espérance μ et une variance σ^2. Posons $\theta = (\mu, \sigma^2)$. Par les méthodes des moments, on obtient le système :

$$\left\{ \begin{array}{l} M_1(\theta) = m_1 \\ \overline{M}_2(\theta) = \overline{m}_2 \end{array} \right. \quad \text{soit} \quad \left\{ \begin{array}{l} \mu = \bar{x}_n \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2 \end{array} \right.$$

d'où $\hat{\mu}_n = \bar{X}_n$ et $\hat{\sigma}_n^2 = V_n^2$.

5.4 Méthode du maximum de vraisemblance : principe

Définition. On appelle *estimateur du maximum de vraisemblance* (EMV) du paramètre θ la statistique $\hat{\theta}_n$ rendant maximale, selon θ, la fonction de vraisemblance du modèle $L(X_1, \ldots, X_n, \theta)$, soit :

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L(X, \theta).$$

Commentaires.

— L'idée de chercher la valeur de θ qui rend maximale la vraisemblance est naturelle : en effet, cette valeur particulière de θ permet de maximiser la probabilité d'obtenir les observations réalisées.

— La technique du maximum de vraisemblance est utilisable quel que soit le modèle utilisé. Notons que si le modèle a de bonnes propriétés de régularité, la détermination de l'EMV $\hat{\theta}_n$ sera simplifiée.

Propriété. Si le modèle vérifie les propriétés **(H1)**, **(H2)** et **(H3)**,
alors $\hat{\theta}_n$ est l'EMV de θ si et seulement si

- $\left.\dfrac{\partial \ln L(X,\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_n} = 0$ soit $S_n(X,\hat{\theta}_n) = 0$ (équation de vraisemblance),
- $\left.\dfrac{\partial^2 \ln L(X,\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}_n} < 0.$

Exemple. Soit le vecteur $(X_1,\ldots,X_n)$ dans les composantes X_i sont i.i.d. de loi $\mathcal{N}(\mu,\sigma^2)$. Posons $\theta = \mu$ et cherchons l'EMV de θ. On a :

- $\left.\dfrac{\partial \ln L}{\partial \theta}\right|_{\theta=\hat{\theta}_n} = \left.\dfrac{\partial \ln L}{\partial \mu}\right|_{\mu=\hat{\mu}_n} = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \hat{\mu}_n) = 0$ (équation de vraisemblance)
  d'où $\hat{\mu}_n = \overline{X}_n$.
- $\left.\dfrac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta=\hat{\theta}_n} = \left.\dfrac{\partial^2 \ln L}{\partial \mu^2}\right|_{\mu=\hat{\mu}_n} = -n/\sigma^2 < 0.$

Donc l'EMV de μ est ici $\hat{\mu}_n$.

Remarque. L'EMV ne peut s'obtenir que si la vraisemblance L est explicite, donc si le modèle est bien explicité (loi normale, loi de Poisson, ...). La méthode des moindres carrés et celle des moments ne nécessitent pas cette spécification.

5.5 Propriétés de l'estimateur du maximum de vraisemblance

Propriété (lien avec l'exhaustivité). S'il existe une statistique exhaustive T_n pour θ, alors l'EMV de θ ne dépend que de T_n.

On donne ci-après, sous forme de théorèmes, trois propriétés asymptotiques de l'EMV $\hat{\theta}_n$. Deux hypothèses supplémentaires sont nécessaires :

(H6) $\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$.
On dit que *le modèle est identifiable*.

(H7) $\dfrac{\partial^2}{\partial \theta^2} f(x,\theta)$ est uniformément continue en x et continue en θ.

Théorème 1. Sous **(H2)** et **(H6)**, on a :

$$\hat{\theta}_n \longrightarrow_{proba} \theta^* \quad \text{quand } n \to +\infty,$$

où la notation θ^* désigne la vraie valeur (inconnue) du paramètre θ. Donc l'EMV $\hat{\theta}_n$ est un estimateur convergent de θ.

Théorème 2. Si les hypothèses **(H1)**, **(H2)**, **(H3)** et **(H4)** sont vérifiées, et si l'équation de vraisemblance admet au moins une solution, alors, avec une probabilité tendant vers 1, la solution est unique et correspond à un maximum.

Théorème 3. Sous les hypothèses **(H1)** à **(H7)**, on a :

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \longrightarrow_{loi} \mathcal{N}(0, I^{-1}(\theta^*)) \quad \text{quand } n \to +\infty.$$

Commentaires.

— La propriété asymptotique donnée au Théorème 2 explique que, dans la pratique, on cherche une valeur annulant l'équation de vraisemblance et on prend cette valeur comme estimation pour $\hat{\theta}_n$. En particulier, dans les cas compliqués (où il n'est pas facile d'obtenir une expression analytique de $\hat{\theta}_n$), les logiciels de statistique utilisent des algorithmes (de type Newton-Raphson) pour obtenir l'estimation du maximum de vraisemblance.

— On peut écrire aussi le résultat du Théorème 3 sous la forme :

$$\sqrt{I_n(\theta^*)}\left(\hat{\theta}_n - \theta^*\right) \longrightarrow_{loi} \mathcal{N}(0,1) \quad \text{quand } n \to +\infty.$$

— Attention, la convergence en loi du Théorème 3 n'entraîne pas $\mathbb{E}[\hat{\theta}_n] \to \theta^*$, ni $n\mathbb{V}(\hat{\theta}_n) \to I^{-1}(\theta^*)$, ni $\mathbb{V}(\hat{\theta}_n) \to I_n^{-1}(\theta^*)$. Pour obtenir ces résultats de convergence, il faut encore rajouter des hypothèses de régularité.

6 Introduction à l'estimation par intervalle de confiance

On va considérer dans ce chapitre un modèle statistique réel paramétrique (avec un paramètre unidimensionnel) :

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta \subset \mathbb{R}))^n.$$

On va construire des intervalles de confiance du paramètre θ. Lorsque θ est multidimensionnel, on parle de régions de confiance.

6.1 Définition

Définition Soit $\alpha \in [0,1]$. On appelle *intervalle de confiance du paramètre* θ de niveau (de confiance) $1-\alpha$ la donnée de deux statistiques A_n et B_n vérifiant

$$P(A_n \leq \theta \leq B_n) = 1 - \alpha.$$

Commentaires.

1. θ représente la vraie valeur (inconnue) du paramètre.
2. A_n et B_n sont deux statistiques réelles, plus précisément il s'agit généralement de deux estimateurs de θ, donc à valeurs dans $(\Theta, \mathcal{B}_\Theta)$.
3. A_n et B_n sont supposées tels que $P(A_n \leq B_n) = 1$.
4. $\alpha \in [0,1]$ est un risque, appelé aussi seuil. La valeur de α est choisie a priori par le statisticien (très souvent, $\alpha = 1\%$, 5% ou 10%).
5. Le niveau de confiance $1-\alpha$ est aussi parfois appelé *coefficient de sécurité* ou *coefficient de confiance*.
6. Soient $x_1, \ldots, x_n$ les valeurs observées des variables aléatoires de l'échantillon $X_1, \ldots, X_n$. Posons $a_n = A_n(x_1, \ldots, x_n)$ et $b_n = B_n(x_1, \ldots, x_n)$. L'intervalle $[a_n, b_n]$ est un intervalle réel inclus dans Θ.
7. Un tel intervalle est aussi parfois appelé *fourchette*, en particulier dans le cadre des sondages.

Remarques

1. Estimer le paramètre θ par intervalle de confiance est plus raisonnable que de l'estimer ponctuellement. En plus de l'intervalle en lui-même, on a une probabilité associée, le niveau de confiance $1-\alpha$.
2. La longueur de l'intervalle de confiance $b_n - a_n$ nous renseigne sur la précision de l'estimation.
3. Il n'existe pas de méthode systématique de construction d'intervalles de confiance. On fait au coup par coup.
4. La notion d'intervalle de confiance est liée à celle de test d'hypothèses, le coefficient α étant alors par exemple le risque de première espèce.

6.2 Liens entre les bornes A_n et B_n

Commençons par remarquer que :

$$P(A_n \leq \theta \leq B_n) = 1-\alpha \quad \Longleftrightarrow \quad \alpha = P(A_n > \theta \text{ ou } B_n < \theta) = P(A_n > \theta) + P(B_n < \theta)$$

On peut être amener à construire des intervalles de confiance de trois types différents.

Cas d'un intervalle du type $[a_n, +\infty[$. Le statisticien cherche ici à assurer une valeur minimale au paramètre θ. C'est par exemple le cas lorsque l'on s'intéresse à la durée de vie minimum d'un composant électronique. On concentre ici le risque α entiérement sur $P(A_n > \theta)$, soit $\alpha = P(A_n > \theta)$. On a en général dans ce cas une solution unique.

Cas d'un intervalle du type $]-\infty, b_n]$. Le statisticien cherche ici à assurer une valeur maximale au paramètre θ. C'est par exemple le cas lorsque l'on désire avoir une concentration en sucre inférieure à un certain pourcentage fixé dans un aliment. On concentre ici le risque α entiérement sur $P(B_n < \theta)$, soit $\alpha = P(B_n < \theta)$. On a en général dans ce cas une solution unique.

Cas d'un intervalle du type $[a_n, b_n]$. Le statisticien cherche ici à encadrer la valeur du paramètre θ. C'est par exemple le cas lorsque l'on s'intéresse au poids d'un paquet de café sur la chaîne de production. On répartit ici le risque α des deux cotés : $\alpha = P(A_n > \theta) + P(B_n < \theta)$.

— Si A_n et B_n ont une distribution symétrique (par exemple, une loi normale), alors on coupe α en $\alpha/2$ et $\alpha/2$, ce qui permet de rendre minimum la longueur $B_n - A_n$. Dans ce cas-là, la solution est unique.

— Si la distribution n'est pas symétrique (par exemple, une loi du χ^2), il n'y a pas de raison de faire comme cela. On a alors une infinité de solutions. Il faut se donner deux valeurs positives α_1 et α_2 telles que $\alpha_1 + \alpha_2 = \alpha$ et poser $\alpha_1 = P(A_n > \theta)$ et $\alpha_2 = P(B_n < \theta)$.

Dans la suite de ce chapitre, on se limitera à présenter des intervalles de confiance du type $[a_n, b_n]$.

7 Introduction aux tests d'hypothèses

Soit un modèle d'échantillonnage $(E, \mathcal{B}, \mathcal{P})^n$ où $(E, \mathcal{B})$ est un espace probabilisable et $\mathcal{P}$ est une famille de loi de probabilités sur $(E, \mathcal{B})$. Ici, on ne considère pas nécessairement

le modèle réel ou paramétrique. Il correspond à l'observation de n variables aléatoires *indépendantes et identiquement distribuées* à valeurs dans $(E, \mathcal{B})$. L'objectif général des tests d'hypothèses est de préciser $\mathcal{P}$.

7.1 Notions d'hypothèses

Définition. On appelle *hypothèse* l'énoncé de toute propriété relative à $\mathcal{P}$. Supposer une telle hypothèse vraie, c'est se restreindre à un sous-ensemble de $\mathcal{P}$.

Exemples.
- Si $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et si $\mathcal{P}$ est la famille des lois continues, on peut faire par exemple l'hypothèse H_0 : *la loi considérée est la loi normale*.
- Si $(E, \mathcal{B}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et si $\mathcal{P}$ est la famille des lois normales, on peut faire par exemple l'hypothèse H_0 : *la moyenne est nulle*.
- Les variables aléatoire X_i considérées sont telles que $X_i = (Y_i, Z_i)$, on peut faire par exemple l'hypothèse H_0 : *Y et Z sont indépendants relativement à $\mathcal{P}$*.

Définitions. L'hypothèse considérée a priori, notée H_0 ci-dessus, est appelée l'*hypothèse nulle*. L'hypothèse prise en compte si H_0 n'est pas retenue est appelée l'*hypothèse alternative* et sera notée H_1.

Remarques.
- Les hypothèses H_0 et H_1 ne jouent pas des rôles symétriques.
- H_0 est une hypothèse considérée comme une hypothèse de travail ; à la limite, à défaut d'information, elle sera retenue. Donc le choix de H_0 est important.

7.2 Notion de test

Un test est une procédure de choix entre les deux hypothèses H_0 et H_1 dans un modèle d'échantillonnage $(E, \mathcal{B}, \mathcal{P})^n$. Tout test sera construit à partir d'une statistique réelle T_n : $(E, \mathcal{B}, \mathcal{P})^n \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Définition. On appelle *test* (ou *test statistique*, ou *test d'hypothèses*) toute règle de décision permettant de choisir entre H_0 et H_1 au vu de T_n.

Dans ce cours, on considérera essentiellement des *tests déterministes* (ou *règles de décision pures*). Plus précisément, on définit une partition de $\mathbb{R}$ en deux classes : $\mathcal{R}_c$ (appelée région critique) et $\mathcal{R}_a$ (appelée région d'acceptation).

La règle de décision est la suivante :
- si $T_n \in \mathcal{R}_c$, on rejette H_0 (on choisit donc H_1).
  On dit que le test est *significatif*.
- Si $T_n \in \mathcal{R}_a$, on ne rejette pas H_0.
  On dit que le test est *non significatif*. On dit qu'on accepte H_0 (par défaut). En fait, rien ne contredit H_0, mais cela ne veut pas dire que cette hypothèse soit "vraie".

7.3 Tests paramétriques, tests non paramétriques

Définition. On appelle *test paramétrique* un test dans lequel les deux hypothèses H_0 et H_1 portent sur les valeurs d'un paramètre, sinon le test est *non paramétrique*.

Exemples.

— Quand les hypothèses sont H_0 : *la loi considérée est la loi normale* et H_1 : *la loi considérée n'est pas la loi normale*, on va faire un test non paramétrique.

— Quand les hypothèses sont H_0 : *la moyenne μ est nulle* et H_1 : $\mu \neq 0$, , on va faire un test paramétrique.

— Quand les hypothèses sont H_0 : *Y et Z sont indépendants relativement à $\mathcal{P}$* et H_1 : *Y et Z ne sont pas indépendants relativement à $\mathcal{P}$*, on va faire un test non paramétrique.

Vocabulaire. Dans un test paramétrique, on appelle *hypothèse simple* une hypothèse de la forme $\theta = \theta_0$ (par exemple $\mu = 0$), sinon on a une *hypothèse composée* : $\theta \neq \theta_0$, $\theta > \theta_0$, $\theta \in [\theta_1, \theta_2]$, ...

7.4 Risque d'erreur dans un test

Faire un test conduit à prendre des décisions dans un "univers aléatoire", cela peut donc conduire à des erreurs.

	Réalité (inconnue) H_0	Réalité (inconnue) H_1
Décision : H_0	bonne décision	erreur de deuxième espèce
Décision : H_1	erreur de première espèce	bonne décision

Définitions.

— On appelle *risque* la probabilité de faire une erreur de décision.

— On appelle *risque de première espèce* la probabilité de rejeter H_0 à tort :

$$\alpha = \mathrm{P}(\text{rejet de } H_0 | H_0 \text{ vraie}).$$

On l'appelle aussi *niveau du test* ou *niveau de signification* ou *seuil*.

— On appelle *risque de deuxième espèce* la probabilité de accepter H_0 à tort :

$$\beta = \mathrm{P}(\text{accepter } H_0 | H_0 \text{ fausse}).$$

— On appelle *puissance d'un test* la probabilité suivante :

$$\pi = 1 - \beta = \mathrm{P}(\text{rejeter } H_0 | H_0 \text{ fausse}).$$

Remarques Le problème est de construire des tests en minimisant les risques associés. En général, l'hypothèse H_0 est simplificatrice, donc il sera souvent plus facile de calculer α que β. Dans la construction d'un test, l'objectif est de minimiser α ou de maximiser π.

Vocabulaire. On dit qu'un test est *convergent* si $\pi_n \to 1$ pour $n \to +\infty$, avec

$$\pi_n = \pi = \mathrm{P}(\text{rejeter } H_0 | H_0 \text{ fausse}) = \mathrm{P}(T_n \in \mathcal{R}_c | H_0 \text{ fausse}).$$

7.5 Quelques rappels utiles sur des lois de probabilité

- $X \sim \mathcal{N}(0,1) \implies X^2 \sim \chi^2(1)$
- $X \sim \mathcal{N}(0,1)$, $Y \sim \chi^2(n)$, X et Y indépendantes $\implies T = \dfrac{X}{\sqrt{Y/n}} \sim T(n)$

- $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$, X_1 et X_2 indépendantes $\implies$ $F = \dfrac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2)$
- $T \sim T(n) \implies F = T^2 \sim F(1, n)$
- *Combinaisons linéaires de variables aléatoires indépendantes* :
  $X_i \sim B(n_i, p)$, X_i indépendantes $\implies$ $\sum_i X_i \sim B(\sum_i n_i, p)$
  $X_i \sim \text{Poisson}(\lambda_i)$, X_i indépendantes $\implies$ $\sum_i X_i \sim \text{Poisson}(\sum_i \lambda_i)$
  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, X_i indépendantes $\implies$ $\sum_i a_i X_i \sim \mathcal{N}(\sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2)$
  $X_i \sim \chi^2(n_i)$, X_i indépendantes $\implies$ $\sum_i X_i \sim \chi^2(\sum_i n_i)$

8 Analyse statistique d'un échantillon

On se place ici dans le cadre où le statisticien dispose d'un seul échantillon $x_1, \ldots, x_n$ à analyser, échantillon qui correspond à une réalisation de $X_1, \ldots, X_n$.

8.1 La théorie sous-jacente

Soient n variables aléatoires réelles $X_1, \ldots, X_n$ indépendantes et de même loi (quelconque), admettant une espérance μ et une variance σ^2. On va tout d'abord se focaliser sur l'estimation des paramètres μ et σ^2 (lorsque l'on suppose que les X_i sont normalement distribuées).

— Un estimateur "naturel" pour μ est $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Cet estimateur est sans biais et on a $\mathbb{V}(\overline{X}_n) = \frac{\sigma^2}{n}$.

— Un estimateur "naturel" pour σ^2 est $V_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$. Cet estimateur est biaisé, mais asymptotiquement sans biais. On peut en déduire un estimateur sans bais de σ^2 : $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

Propriétés fondamentales des statistiques $\overline{X}_n$ et S_n^2. On rappelle que si l'on suppose que chaque variable aléatoire réelle X_i suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, alors on a :

— $\overline{X}_n \sim \mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$.

— $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1)$.

— Les statistiques $\overline{X}_n$ et S_n^2 sont indépendantes.

8.2 Intervalles de confiance pour la moyenne μ et pour la variance σ^2.

On suppose ici que l'on dispose d'un échantillon $(X_1, \ldots, X_n)$ où les X_i sont indépendants et identiquement distribué selon la loi $\mathcal{N}(\mu, \sigma^2)$. On s'intéressera successivement à la moyenne μ et à la variance σ^2. On travaille avec un niveau de confiance $(1-\alpha)$ fixé.

Définition Une fonction pivotale f pour le paramètre θ est une fonction de $(X_1, \ldots, X_n)$ et du paramètre θ telle que la loi de $f(X_1, \ldots, X_n, \theta)$ ne dépend pas du paramètre θ.

— **Intervalle de confiance pour μ lorsque σ^2 est *connue*.**
On utilise la fonction pivotale suivante :

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

L'intervalle de confiance pour μ de niveau de confiance $1-\alpha$ lorsque σ^2 est connue est alors :

$$\overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

où $z_{1-\alpha/2}$ est le fractile d'ordre $1-\alpha/2$ de la loi normale centrée réduite $\mathcal{N}(0,1)$.

— **Intervalle de confiance pour μ lorsque σ^2 est** *inconnue*.
On ne peut plus utiliser la fonction pivotale Z_n car on ne connait pas σ. On va alors utiliser la fonction pivotale suivante :

$$T_n = \frac{\overline{X}_n - \mu}{S_n} \sim T(n-1).$$

L'intervalle de confiance pour μ de niveau de confiance $1-\alpha$ lorsque σ^2 est inconnue est alors :

$$\overline{X}_n - t_{1-\alpha/2}\frac{S_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + t_{1-\alpha/2}\frac{S_n}{\sqrt{n}}$$

où $t_{1-\alpha/2}$ est le fractile d'ordre $1-\alpha/2$ de la loi de Student $T(n-1)$ et $S_n = \sqrt{S_n^2}$.

— **Intervalle de confiance pour σ^2 lorsque μ est** *connue*.
On se donne ici $\alpha_1 > 0$ et $\alpha_2 > 0$ vérifiant $\alpha_1 + \alpha_2 = \alpha$. On utilise ici le fait que les variables $Z_i = \frac{X_i - \mu}{\sigma}$ sont indépendantes et de même loi $\mathcal{N}(0,1)$; on en déduit alors que la fonction pivotale suivante :

$$\widetilde{K}_n = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n).$$

L'intervalle de confiance pour σ^2 de niveau de confiance $1-\alpha$ lorsque μ est connue est alors :

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\tilde{k}_2} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\tilde{k}_1}$$

où $\tilde{k}_1$ (resp. $\tilde{k}_2$) est le fractile d'ordre α_1 (resp. $1-\alpha_2$) de la loi du chi-deux $\chi^2(n)$.

— **Intervalle de confiance pour σ^2 lorsque μ est** *inconnue*.
On se donne ici à nouveau $\alpha_1 > 0$ et $\alpha_2 > 0$ vérifiant $\alpha_1 + \alpha_2 = \alpha$. On utilise ici la fonction pivotale suivante :

$$K_n = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1).$$

L'intervalle de confiance pour σ^2 de niveau de confiance $1-\alpha$ lorsque μ est inconnue est alors :

$$\frac{\sum_{i=1}^{n}(X_i-\overline{X}_n)^2}{k_2} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-\overline{X}_n)^2}{k_1}$$

où k_1 (resp. k_2) est le fractile d'ordre α_1 (resp. $1-\alpha_2$) de la loi du chi-deux $\chi^2(n-1)$.

Remarques.

— En général, les cas où σ^2 (ou μ) est connue sont rares. Les deux autres cas sont les plus usuels en pratique.

— Les quantités (bornes) intervenant dans les intervalles de confiance de la variance σ^2 sont strictement positives, on peut donc en déduire un intervalle de confiance pour l'écart-type σ au niveau de confiance $1-\alpha$: par exemple, lorsque l'on suppose μ connue, on a

$$\sqrt{\frac{\sum_{i=1}^n (X_i-\mu)^2}{\tilde{k}_2}} \leq \sigma \leq \sqrt{\frac{\sum_{i=1}^n (X_i-\mu)^2}{\tilde{k}_1}}.$$

8.3 Tests d'hypothèses portant sur la moyenne μ ou la variance σ^2.

Soient n variables aléatoires $X_1, \ldots, X_n$ indépendantes et identiquement distribuées selon la loi normale $\mathcal{N}(\mu, \sigma^2)$. On cherche à faire des tests d'hypothèses portant successivement sur la moyenne μ ou la variance σ^2.

— **Test portant sur la moyenne μ (test de Student)**.
Pour tester l'hypothèse nulle $H_0 : \mu = \mu_0$ contre une des hypothèses alternatives suivantes $H_1 : \mu \neq \mu_0$ ou $H_1 : \mu > \mu_0$ ou $H_1 : \mu < \mu_0$. Pour cela, on considère la statistique

$$T_n = \frac{\overline{X}_n - \mu_0}{S_n}.$$

Cette statistique est calculable car, à partir de l'échantillon $X_1, \ldots, X_n$, on peut calculer $\overline{X}_n$ et S_n, et on se donne la valeur μ_0 quand on définit notre hypothèse. Par contre, on ne connait la loi de T_n que lorsque l'on se place sous l'hypothèse H_0 : en effet dans ce cas-là, la vraie moyenne μ des X_i est μ_0. On a alors :

$$\text{sous } H_0, \quad T_n = \frac{\overline{X}_n - \mu_0}{S_n} \sim T(n-1).$$

En prenant un risque de première espèce α, on peut définir les régions de rejet associées à chaque des hypothèses alternatives : on rejettera l'hypothèse H_0 si $T_n \in R_\alpha$ avec

$$\begin{array}{ll} \text{pour } H_1 : \mu \neq \mu_0, & R_\alpha =]-\infty, t_{\alpha/2}(n-1)] \cup [t_{1-\alpha/2}(n-1), +\infty[, \\ \text{pour } H_1 : \mu > \mu_0, & R_\alpha = [t_{1-\alpha}(n-1), +\infty[, \\ \text{pour } H_1 : \mu < \mu_0, & R_\alpha =]-\infty, t_\alpha(n-1)]. \end{array}$$

— **Test portant sur la variance σ^2**.
Pour tester l'hypothèse nulle $H_0 : \sigma^2 = \sigma_0^2$ contre une des hypothèses alternatives suivantes $H_1 : \sigma^2 \neq \sigma_0^2$ ou $H_1 : \sigma^2 > \sigma_0^2$ ou $H_1 : \sigma^2 < \sigma_0^2$. Pour cela, on considère la statistique

$$K_n = \frac{(n-1)S_n^2}{\sigma_0^2}.$$

Cette statistique est calculable car, à partir de l'échantillon $X_1, \ldots, X_n$, on peut calculer S_n^2, et on se donne la valeur σ_0^2 quand on définit notre hypothèse. Par contre, on ne connait la loi de K_n que lorsque l'on se place sous l'hypothèse H_0 : en effet dans ce cas-là, la vraie variance σ^2 des X_i est σ_0^2. On a alors :

$$\text{sous } H_0, \quad K_n = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2(n-1).$$

En prenant un risque de première espèce α, on peut définir les régions de rejet associées à chaque des hypothèses alternatives : on rejettera l'hypothèse H_0 si $K_n \in R_\alpha$ avec

$$\begin{array}{ll} \text{pour } H_1 : \sigma^2 \neq \sigma_0^2, & R_\alpha =]-\infty, \chi^2_{\alpha/2}(n-1)] \cup [\chi^2_{1-\alpha/2}, +\infty[, \\ \text{pour } H_1 : \sigma^2 > \sigma_0^2, & R_\alpha = [\chi^2_{1-\alpha}, +\infty[, \\ \text{pour } H_1 : \sigma^2 < \sigma_0^2, & R_\alpha =]-\infty, \chi^2_{\alpha}(n-1)]. \end{array}$$

9 Analyse statistique de deux échantillons

On se place dans le cadre où le statisticien dispose de deux échantillons à analyser. On va considérer deux cas :

- cas de deux échantillons appariés,
- cas de deux échantillons indépendants.

9.1 Un peu de théorie : cas de deux échantillons appariés

On dispose ici de n couples $\{(X_i, Y_i),\ i = 1, \ldots, n\}$ indépendants et identiquement distribués avec $\mathbb{E}[X_i] = \mu_1$ et $\mathbb{E}[Y_i] = \mu_2$. On désire tester l'hypothèse :

$$H_0 \ : \ \mu_1 = \mu_2.$$

Remarque. Les couples (X_i, Y_i) sont indépendants, mais les variables X_i et Y_i sont appariées. Par exemple, elles peuvent modéliser la taille du père et celle du fils, ou bien la tension artérielle avant un effort et la tension artérielle après un effort, ...

Démarche théorique. On pose $Z_i = Y_i - X_i$: les variables Z_i sont indépendants et indentiquement distribuées de moyenne $\mathbb{E}[Z_i] = \mu = \mu_1 - \mu_2$. Ainsi tester $H_0 \ : \ \mu_1 = \mu_2$ revient à tester l'hypothèse $H_0 \ : \ \mu = 0$. Si l'on peut supposer la normalité des Z_i, on est alors ramené au cas précédent où l'on a un seul échantillon (celui des différences) dont on veut tester la nullité de la moyenne.

Remarque. La normalité des X_i et celle des Y_i ne sont ni nécessaires, ni suffisantes. Seule la normalité des différences Z_i est nécessaire. Lorsque $n \geq 30$ et même si les Z_i ne sont pas normalement distribuées, on peut utiliser ce test qui est robuste. Sinon, on doit utiliser des tests non paramètriques (ne reposant pas sur des hypothèses de normalité).

9.2 Un peu de théorie : cas de deux échantillons indépendants

On dispose ici deux échantillons :

— Echantillon 1 : $X_1, \ldots, X_{n_1}$ composé de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{N}(\mu_1, \sigma_1^2)$;

— Echantillon 2 : $Y_1, \ldots, Y_{n_2}$ composé de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{N}(\mu_2, \sigma_2^2)$;

— Les deux échantillons sont globalement indépendants.

On considère les statistiques suivantes :

$$\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \ \ S_X^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i \overline{X})^2, \ \ \overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i, \ \ S_Y^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i \overline{Y})^2.$$

— **Test d'égalité des variances (test de Fisher).**

On teste ici l'hypothèse $H_0 \ : \ \sigma_1^2 = \sigma_2^2$. On utilise la statistique suivante

$$F_n = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

En effet, on a $K_1 = \frac{(n_1-1)S_X^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$, $K_2 = \frac{(n_2-1)S_Y^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$, et ces deux quantités sont indépendantes car elles sont issues d'échantillons indépendants. On peut alors en déduire la loi de $F_n = \frac{K_1/(n_1-1)}{K_2/(n_2-1)}$. Sous H_0, cette statistique est calculable et vaut $F_n = \frac{S_X^2}{S_Y^2}$.

Remarque : on peut aussi considérer la statistique $F_n^* = \frac{S_Y^2}{S_X^2} \sim F(n_2 - 1, n_1 - 1)$ sous H_0.

• Cas où l'hypothèse alternative est $H_1 \ : \ \sigma_1^2 \neq \sigma_2^2$:
Soient α_1 et α_2 tels que $\alpha_1 + \alpha_2 = \alpha$ (souvent on pose $\alpha_1 = \alpha_2 = \alpha/2$. La région critique est
soit $\mathcal{R}_c : F_n < F_{\alpha_1}(n_1 - 1, n_2 - 1)$ ou $F_n > F_{1-\alpha_2}(n_1 - 1, n_2 - 1)$;
soit $\mathcal{R}_c^* : F_n^* < F_{\alpha_1}(n_2 - 1, n_1 - 1)$ ou $F_n^* > F_{1-\alpha_2}(n_2 - 1, n_1 - 1)$.
Ces deux régions critiques sont équivalente lorsque $\alpha_1 = \alpha_2 = \alpha/2$. Dans ce cas-là, on peut écrire la règle de décision sous la forme :
$\mathcal{R}_c : \frac{\sup(S_X^2, S_Y^2)}{\inf(S_X^2, S_Y^2)} > F_{1-\alpha/2}(\nu_1, \nu_2)$
avec $\nu_1 = \begin{cases} n_1 - 1 & \text{si } \sup(S_X^2, S_Y^2) = S_X^2 \\ n_2 - 1 & \text{sinon.} \end{cases}$ et $\nu_2 = \begin{cases} n_2 - 1 & \text{si } \nu_1 = n_1 - 1 \\ n_1 - 1 & \text{sinon.} \end{cases}$

• Cas où l'hypothèse alternative est $H_1 \ : \ \sigma_1^2 < \sigma_2^2$:
La région critique est
soit $\mathcal{R}_c : F_n < F_\alpha(n_1 - 1, n_2 - 1)$;
soit $\mathcal{R}_c^* : F_n^* > F_{1-\alpha}(n_2 - 1, n_1 - 1)$.
Ces deux règles de décision sont équivalentes. Par convention, on choisit la règle de décision basée sur le fractile d'ordre $1 - \alpha$ de la loi de Fisher sous-jacente.

• Cas où l'hypothèse alternative est $H_1 \ : \ \sigma_1^2 > \sigma_2^2$:
La région critique est
soit $\mathcal{R}_c : F_n > F_{1-\alpha}(n_1 - 1, n_2 - 1)$;
soit $\mathcal{R}_c^* : F_n^* < F_\alpha(n_2 - 1, n_1 - 1)$.
Ces deux règles de décision sont elles-aussi équivalentes. Par convention, on choisit la règle de décision basée sur le fractile d'ordre $1 - \alpha$ de la loi de Fisher sous-jacente.

— **Test d'égalité des moyennes quand les variances sont supposées égales.**
On suppose que $\sigma_1^2 = \sigma_2^2 = \sigma^2$. On veut tester $H_0 \ : \ \mu_1 = \mu_2$. On a : $\overline{X} \sim \mathcal{N}(\mu_1, \sigma^2/n_1)$, $\overline{Y} \sim \mathcal{N}(\mu_2, \sigma^2/n_2)$, $\overline{X}$ et $\overline{Y}$ sont indépendantes. On en déduit que :

$$\overline{X} - \overline{Y} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$$

et donc que :

$$U_n = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{N}(0, 1).$$

De plus, on a : $\frac{(n_1-1)S_X^2}{\sigma^2} \sim \chi^2(n_1 - 1)$, $\frac{(n_2-1)S_Y^2}{\sigma^2} \sim \chi^2(n_2 - 1)$, et ces deux quantités

sont indépendantes. On en déduit que :

$$\frac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{\sigma^2} \sim \chi^2(n_1+n_2-2).$$

On peut montrer que le meilleur estimateur de la variance commune σ^2 est

$$S^2 = \frac{1}{n_1+n_2-2}\left(\sum_{1=1}^{n_1}(X_i-\overline{X})^2+\sum_{1=1}^{n_2}(Y_i-\overline{Y})^2\right).$$

Il est facile de voir que

$$W_n = \frac{(n_1+n_2-2)S^2}{\sigma^2} \sim \chi^2(n_1+n_2-2).$$

Enfin, on peut montrer que les statistique U_n et W_n sont indépendantes.
Finalement, on obtient donc la statistique

$$T_n = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_n\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim T(n_1+n_2-2).$$

Sous l'hypothèse H_0, cette statistique est calculable :

$$T_n = \frac{\overline{X}-\overline{Y}}{S_n\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim T(n_1+n_2-2).$$

• Cas où l'hypothèse alternative est $H_1 \ : \ \mu_1 \neq \mu_2$:
La région critique est $\mathcal{R}_c : |T_n| > T_{1-\alpha/2}(n_1+n_2-2)$.
• Cas où l'hypothèse alternative est $H_1 \ : \ \mu_1 > \mu_2$:
La région critique est $\mathcal{R}_c : T_n > T_{1-\alpha}(n_1+n_2-2)$.
• Cas où l'hypothèse alternative est $H_1 \ : \ \mu_1 < \mu_2$:
La région critique est $\mathcal{R}_c : T_n < -T_{1-\alpha}(n_1+n_2-2)$.

— **Test d'égalité des moyennes quand les variances sont supposées inégales.**
On veut toujours tester $H_0 \ : \ \mu_1 =_m u_2$. On va faire un test approché (asymptotique). On a : $\overline{X} \sim \mathcal{N}(\mu_1, \sigma_1^2/n_1)$, $\overline{Y} \sim \mathcal{N}(\mu_2, \sigma_2^2/n_2)$, $\overline{X}$ et $\overline{Y}$ sont indépendantes. On en déduit que :

$$\overline{X}-\overline{Y} \sim \mathcal{N}(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$

et donc que :

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0,1).$$

On utilise le fait que : $S_X^2 \to \sigma_1^2$ en probabilité pour $n_1 \to +\infty$, et $S_Y^2 \to \sigma_2^2$ en probabilité pour $n_2 \to +\infty$. On a alors, pour $n_1 \to +\infty$ et $n_2 \to +\infty$, le résultat de convergence suivant :

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_X^2}{n_1}+\frac{S_Y^2}{n_2}}} \sim \mathcal{N}(0,1).$$

Pour faire le test, on utilise alors la statisque suivante :

$$\text{sous } H_0, \quad Z_n = \frac{(\overline{X} - \overline{Y})}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}} \sim \mathcal{N}(0,1).$$

Pour $H_1 \ : \ \mu_1 \neq \mu_2$, la région critique est $\mathcal{R}_c : |Z_n| > z_{1-\alpha/2}$.
Pour $H_1 \ : \ \mu_1 > \mu_2$, la région critique est $\mathcal{R}_c : Z_n > z_{1-\alpha}$.
Pour $H_1 \ : \ \mu_1 < \mu_2$, la région critique est $\mathcal{R}_c : Z_n < -z_{1-\alpha}$.

Remarque Si les échantillons sont normalement distribués, cette approche est valable pour n_1 et n_2 assez grands ($n_1 + n_2 \geq 40$, $n_1 \geq 15$, $n_2 \geq 15$). Si les échantillons ne proviennent pas de distributions normales, on peut encore utiliser ce test pour $n_1 + n_2 \geq 100$, $n_1 \geq 30$, $n_2 \geq 30$. En dehors de ces cas, il est préférable d'utiliser des tests non paramétriques.

Table de la loi du χ^2

$$\Pr[X \leq x] = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}}\, y^{r/2-1} e^{-r/2}\, dx$$

	$\Pr[X \leq x]$					
r	0.01	0.025	0.05	0.95	0.975	0.99
1	0.000	0.001	0.004	3.841	5.024	6.635
2	0.020	0.051	0.103	5.991	7.378	9.210
3	0.115	0.216	0.352	7.815	9.348	11.345
4	0.297	0.484	0.711	9.488	11.143	13.277
5	0.554	0.831	1.145	11.070	12.833	15.086
6	0.872	1.237	1.635	12.592	14.449	16.812
7	1.239	1.690	2.167	14.067	16.013	18.475
8	1.646	2.180	2.733	15.507	17.535	20.090
9	2.088	2.700	3.325	16.919	19.023	21.666
10	2.558	3.247	3.940	18.307	20.483	23.209
11	3.053	3.816	4.575	19.675	21.920	24.725
12	3.571	4.404	5.226	21.026	23.337	26.217
13	4.107	5.009	5.892	22.362	24.736	27.688
14	4.660	5.629	6.571	23.685	26.119	29.141
15	5.229	6.262	7.261	24.996	27.488	30.578
16	5.812	6.908	7.962	26.296	28.845	32.000
17	6.408	7.564	8.672	27.587	30.191	33.409
18	7.015	8.231	9.390	28.869	31.526	34.805
19	7.633	8.907	10.117	30.144	32.852	36.191
20	8.260	9.591	10.851	31.410	34.170	37.566
21	8.897	10.283	11.591	32.671	35.479	38.932
22	9.542	10.982	12.338	33.924	36.781	40.289
23	10.196	11.689	13.091	35.172	38.076	41.638
24	10.856	12.401	13.848	36.415	39.364	42.980
25	11.524	13.120	14.611	37.652	40.646	44.314
26	12.198	13.844	15.379	38.885	41.923	45.642
27	12.879	14.573	16.151	40.113	43.195	46.963
28	13.565	15.308	16.928	41.337	44.461	48.278
29	14.256	16.047	17.708	42.557	45.722	49.588
30	14.953	16.791	18.493	43.773	46.979	50.892

Table de la loi de Student

$$\Pr[T \leq t] = \int_{-\infty}^{t} \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \frac{1}{(1+x^2/r)^{(r+1)/2}}\, dx$$

$$\Pr[T \leq -t] = 1 - \Pr[T \leq t]$$

	$\Pr[T \leq t]$				
r	0.90	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750

Table de la loi de Fisher

$$\Pr[F_{r_1,r_2} \leq f] = \int_0^f \frac{\Gamma((r_1+r_2)/2)(r_1/r_2)^{r_1/2}x^{(r_1/2-1)}}{\Gamma(r_1/2)\Gamma(r_2/2)(1+r_1x/r_2)^{(r_1+r_2)/2}}\,dx$$

		r_1											
$\Pr[F \leq f]$	r_2	1	2	3	4	5	6	7	8	9	10	12	15
0.95	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95
0.975		647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	976.71	984.87
0.99		4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.28
0.95	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43
0.975		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43
0.99		98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43
0.95	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
0.975		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25
0.99		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87
0.95	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
0.975		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66
0.99		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20
0.95	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
0.975		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43
0.99		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72
0.95	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
0.975		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27
0.99		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56

		r_1											
$\Pr[F \leq f]$	r_2	1	2	3	4	5	6	7	8	9	10	12	15
0.95	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
0.975		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57
0.99		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31
0.95	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
0.975		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10
0.99		11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52
0.95	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
0.975		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77
0.99		10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96
0.95	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
0.975		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52
0.99		10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56
0.95	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
0.975		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18
0.99		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01
0.95	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
0.975		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86
0.99		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52