Calcul des probabilités & théorie des erreurs.

Xavier Dubuc

24 octobre 2009

Table des matières

Chapitre 0 - Introduction

Vers le 1er décembre, un problème sera posé de manière individuelle et chaque étudiant devra répondre à son problème au travers d'un rapport (entre 20 et 50 pages) qu'il remettra à l'enseignant avant les vacances de noël. Le style est libre mais il convient, comme dans tous rapports, de constituer celui-ci d'une introduction, d'un corps, d'une conclusion et d'éventuellement des annexes. Il est également inutile de recopier les démonstrations du cours, un simple «comme vu au cours à tel endroit» convient ; cependant, l'étudiant doit être capable d'expliquer la théorie cernée par cette phrase.

Au mois de janvier, une défense orale durant entre 10 et 30 minutes aura lieu. Au cours de celle-ci, l'étudiant sera amené à expliquer son rapport dans le but de convaincre le jury que ce rapport a bien été fait par celui-ci. Il est à savoir que le professeur peut demander à argumenter le choix de la méthode de résolution utilisée et ce en demandant des explications sur une partie du cours n'étant pas spécialement utilisée dans le rapport de l'étudiant. Il convient donc de revoir tout le cours dans son entiereté, de le comprendre et surtout d'être capable de l'appliquer.

Au niveau du support de cours, un syllabus est disponible au presse, son intitulé est le même que celui du cours. Ce dernier contient la théorie inhérente au cours ainsi que des exercices et des résolutions d'exercices. Il contient également des tables statistiques dont l'étudiant aura usage lors des séances d'exercices ainsi que 2 rapports d'anciens élèves jugés bons à titre d'exemple pour les étudiants.

Il est conseillé de lire sa question dès sa réception car le problème posé peut être dur à cerner, il ne faut dès lors pas attendre la dernière minute pour oser poser une question par rapport au problème. A cet effet, dans les environs du 8 décembre, aura lieu une séance spéciale d'exercices durant laquelle tous les étudiants pourront poser des questions par rapport à leur problème ainsi qu'assister à la résolution d'un exercice similaire.

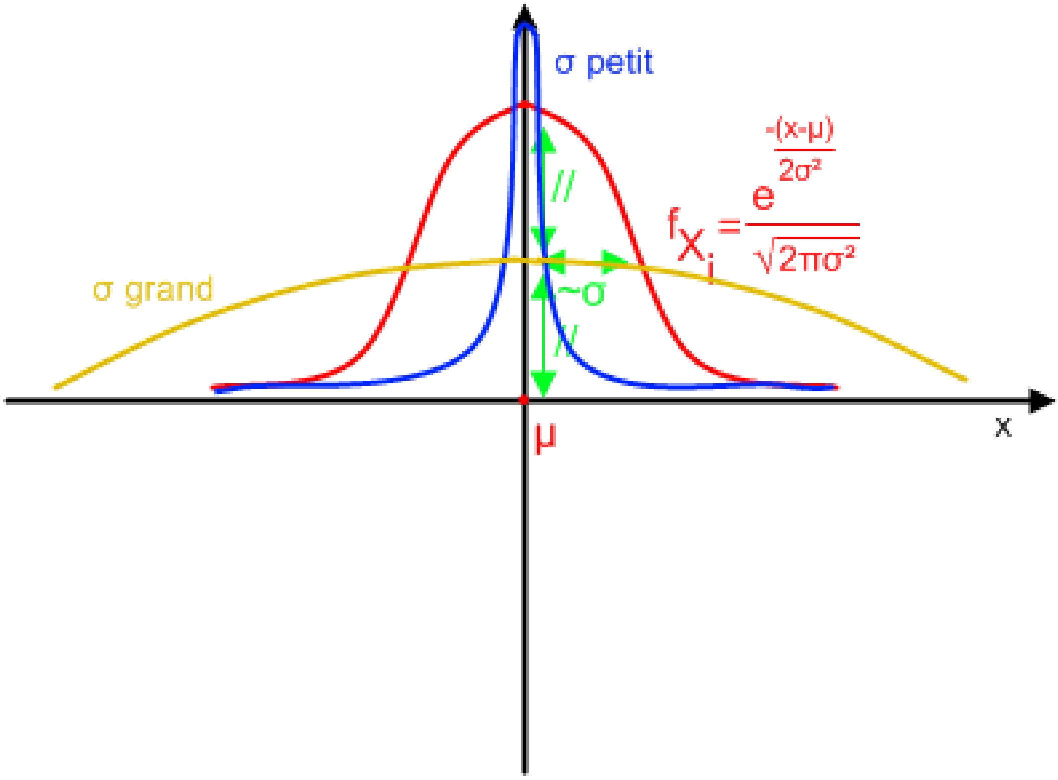
1 Chapitre 1 - 4 lois de référence

1.1 Loi Normale

Prenons, par exemple, X_i comme étant le temps nécessaire à produire une voiture. On a que le temps est égal à la somme de tous les temps pour mettre chaque pièce sur la voiture (une voiture étant composée de plus ou moins 5000 pièces). On a donc : temps $= \sum_{i=1}^{N} T_i$. Vu $N >> 1$ (N (= 5000) est beaucoup plus grand que 1), le **théorème de la limite centrale** nous dit que le temps sera de **loi normale**.

On pose :
- X_1 v.a.r. associée au temps d'assemblage de la 1ère voiture : mesure x_1.
- X_2 v.a.r. associée au temps d'assemblage de la 2ème voiture : mesure x_2.
- ...
- X_i v.a.r. associée au temps d'assemblage de la ième voiture : mesure x_i.
- ...
- X_n v.a.r. associée au temps d'assemblage de la nème voiture : mesure x_n.

X_i est de loi normale, $\rightarrow f_{X_i} = \frac{e^{\frac{-(x-\mu)}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$



La courbe est centrée en μ qui est la moyenne arithmétique, et la largeur (distance entre l'axe vertical passant par μ et la courbe) à mi-hauteur est, à peu de choses près, égale à σ. On déduit les formes des courbes sur le graphe ci-haut par le fait que vu que c'est une loi de probabilité, il faut vérifier que $\int f(x)\ fx = 1$ et on remarque également que plus σ est petit, plus la densité est forte autour de la moyenne et plus les mesures sont précises. On va donc toujours essayer d'avoir un σ très petit.

Interessons nous à présent à la loi de la moyenne algébrique des v.a.r. : $\bar{X}_N = \frac{X_1 + X_2 + ... + X_N}{N}$, on cherche donc $f_{\bar{X}_N}(x) = ?$. Il existe plusieurs façons de le calculer, la plus facile reste d'utiliser les fonctions caractéristiques. Pour X_j, la fonction caractéristique est donnée par : $\varphi_{X_j}(t) = E\{e^{itX_j}\} = e^{it\mu - \sigma^2 \frac{t^2}{2}}$.

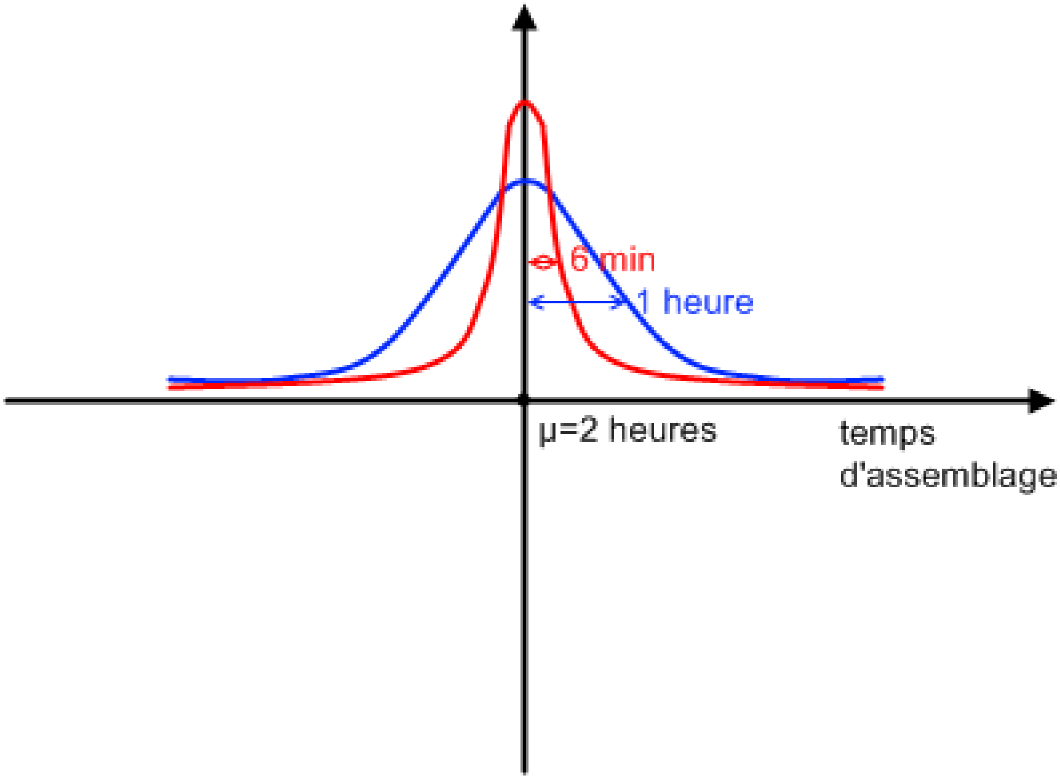
En supposant que les X_j sont tous indépendants et vu qu'ils sont tous de même loi, on peut déduire les égalités suivantes :

$\varphi_{\bar{X}_N}(t)$
$= E\{e^{i}t\frac{X_1+...+X_N}{N}\}$
$= E\{e^{i\frac{t}{N}X_1+...+X_N}\}$
$= E\{e^{i\frac{t}{N}X_1}\} * E\{e^{i\frac{t}{N}X_2}\}...E\{e^{i\frac{t}{N}X_N}\}$

$= \varphi_{X_1}\left(\frac{t}{N}\right)...\varphi_{X_N}\left(\frac{t}{N}\right)$
$= \left(\varphi_{X_1}\left(\frac{t}{N}\right)\right)^N$
$= \left(e^{i\frac{t}{N}\mu-\sigma^2\frac{t^2}{2N^2}}\right)^N$
$= e^{i\frac{t}{N}\mu N-\sigma^2\frac{t^2}{2N^2}N}$
$= e^{it\mu-\sigma^2\frac{t^2}{2N}}$
$= e^{it\mu-\frac{\sigma^2}{N}\frac{t^2}{2}}$

Nous étions entrain de considérer une loi normale de moyenne μ et d'écart-type σ, $f_{X_i}(x) = N(\mu, \sigma)$
$\Rightarrow \varphi_{X_j}(t) = e^{it\mu-\sigma^2\frac{t^2}{2}}$
$\Rightarrow \varphi_{\bar{X}_N}(t) = e^{it\mu-\frac{\sigma^2}{N}\frac{t^2}{2}} = N(\mu, \frac{\sigma}{\sqrt{N}})$

Il s'agit donc d'une gaussienne dont l'écart-type (la largeur à mi-hauteur) est plus petit, on peut donc en conclure que lorsque l'on considère la moyenne d'une suite d'évènements indépendants on augmente la précision des mesures. Revenons à l'exemple de la construction de la voiture :



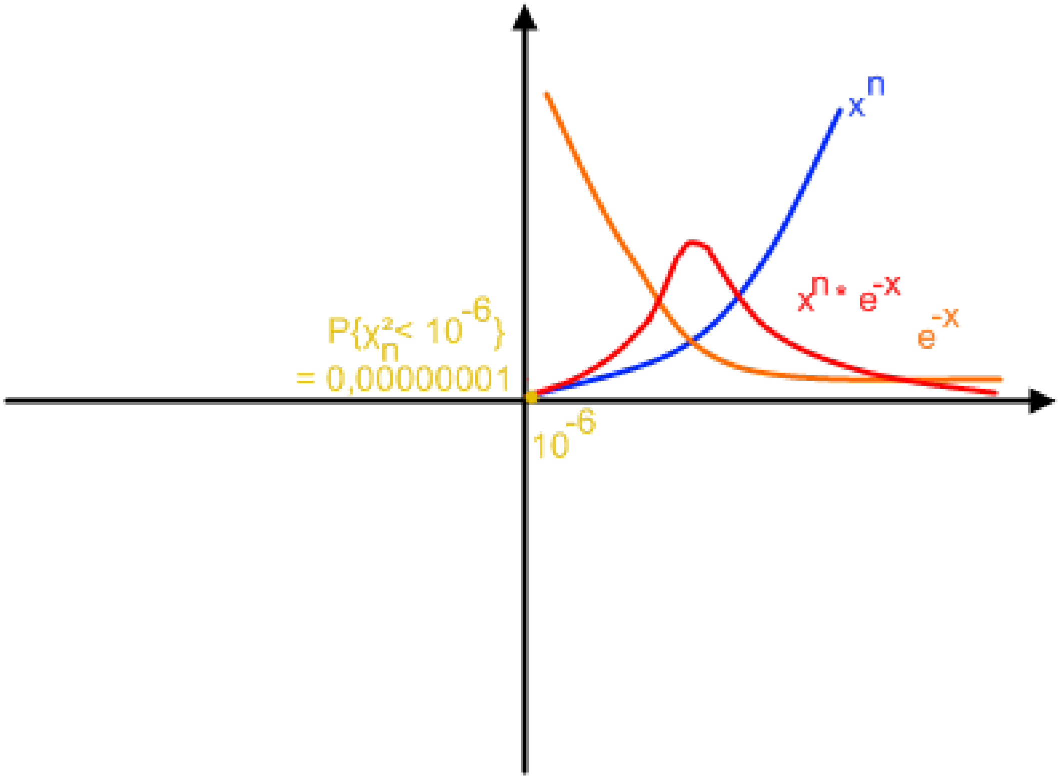
$\bar{X}_{100} = \frac{X_1 + ... + X_{100}}{100}$ $(6\text{min} = \frac{1h}{\sqrt{100}})$

1.2 Loi Chi 2 à n degrés de liberté

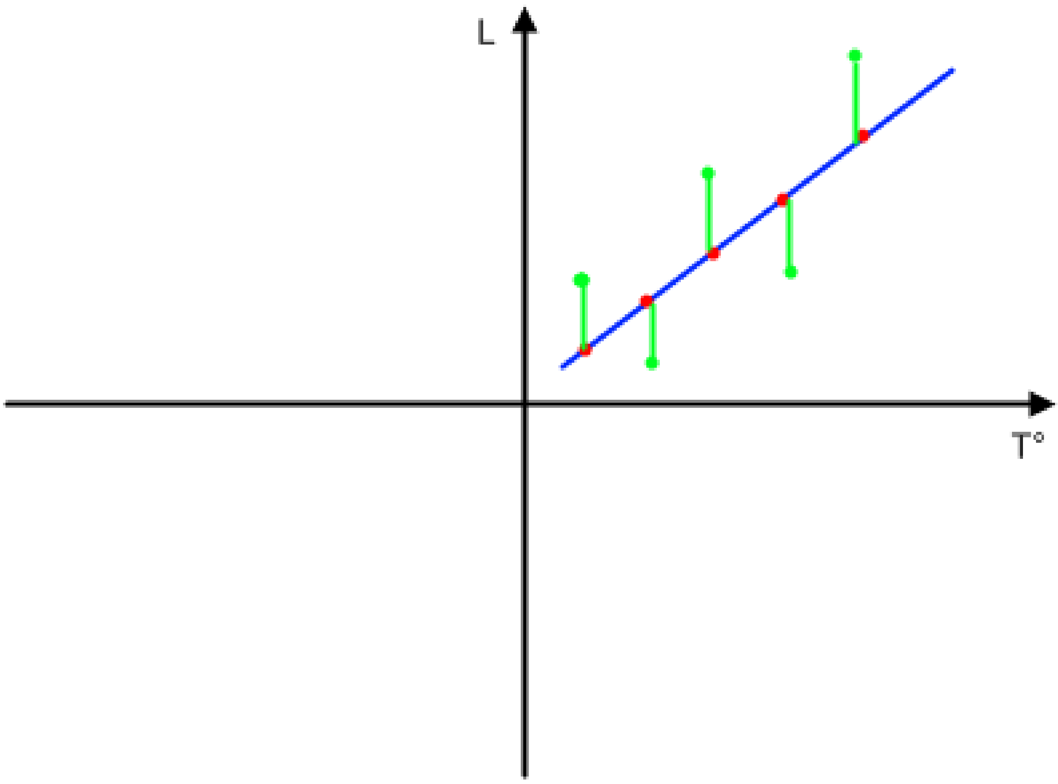
Prenons $X_i = N(0,1)$, la loi normale centrée réduite (moyenne nulle et variance 1). Il existe un moyen de «transformer» une loi normale en une loi normale centrée réduite, imaginons que Y soit une loi normale, alors $\frac{Y-\mu}{\sigma}$ est une loi normale centrée réduite.

On considère ensuite les «Chi 2», il s'agit en fait de la somme de toutes les mesures au carré : $\chi_n^2 = \sum_{i=1}^n X_i^2$. On donne également sa fonction de répartition : $f_{\chi_n^2}(x) = \begin{cases} \frac{1}{2\Gamma(\frac{n}{2})}(\frac{x}{2})^{\frac{n}{2}-1}e^{-\frac{x}{2}} \; si \; x > 0 \\ 0 \; ailleurs \end{cases}$
Essayons de deviner l'aspect de la courbe de $f_{\chi_n^2}$, son équation est du type : $Cste * (\frac{x}{2})^{\frac{n}{2}-1}e^{-\frac{x}{2}}$; on

peut ainsi deviner que si x est petit, on a $f_{\chi_n^2} \sim (\frac{x}{2})^{\frac{n}{2}-1}$ et qu'au contraire, quand x est grand, on a $f_{\chi_n^2} \sim e^{-\frac{x}{2}}$. On en déduit le graphe suivant :



Exemple : lors d'un labo de première année, les premières physiques ont été amené à mesurer la dilatation d'un tube de cuivre en fonction de la température, ils ont eu une série de mesure à faire et par la suite le prof leur a dit que les points trouvés formaient une courbe linéaire. Dès lors les élèves ont modifiés leurs graphiques afin que les points tombent sur la droite du professeur. Cette droite est en fait la moyenne des mesures faites, et on va montrer avec l'exemple suivant que la probabilité pour que tous les points (ici 5) se trouvent sur cette droite est infime voire nulle.



$\sum_i \frac{(X_i - (aT+b))}{\sigma^2} = \chi_n^2$

$\Rightarrow P\{\chi_n^2 < 10^{-6}\} = ?$

Suivons l'exemple avec $n = 5$, $P\{\chi_5^2 < 10^{-6}\} = P\{0 \leq \chi_5^2 \leq 10^{-6}\} = 0,00000001$, ce qui représente la probabilité que tous les points soient sur la droite.

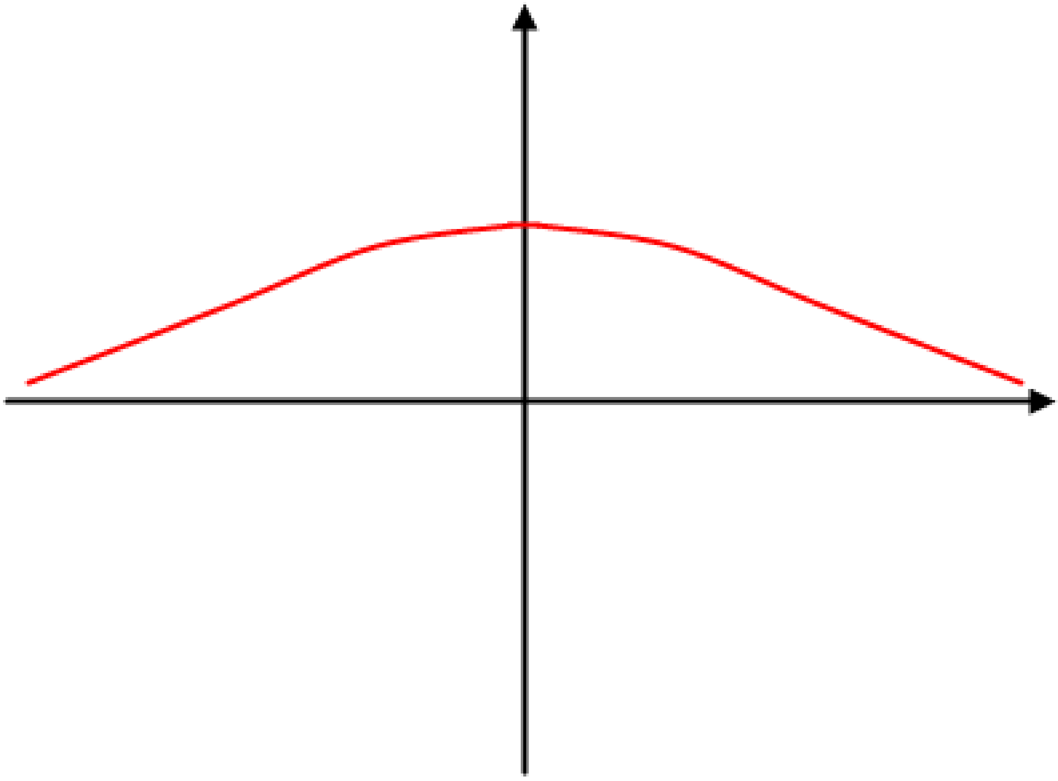
La réponse la plus probable, c'est quand χ_n^2 est maximum, c'est à dire aux alentours de n, on dit familièrement que χ_n^2 «aime» les données qui sont de l'ordre du degré de liberté.

1.3 Loi Student à n degrés de liberté

On prend :

$\begin{cases} X = N(0,1) \\ Y = \chi_n^2 \end{cases}$ avec X et Y indépendants.

On définit la loi de **Student** à n degré de liberté par : $T_n = S_n = \dfrac{X}{\sqrt{\frac{Y}{n}}}$ dont la fonction de répartition est $f_{S_n}(t) = Cste \dfrac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}}$



1.4 Loi Fisher-Snedecor à n_1,n_2 degrés de liberté

On prend :
$\begin{cases} X_1 = \chi^2_{n_1} \\ X_2 = \chi^2_{n_2} \end{cases}$ avec X_1 et X_2 indépendants.

On définit la loi de **Fischer-Snedecor** à n degré de liberté par : $F_{n_1,n_2} = \frac{\frac{X_1}{n_1}}{\frac{X_2}{n_2}}$ dont la fonction de répartition est $f(u) = \begin{cases} Cte\dfrac{u^{\frac{n_1}{2}}-1}{(1+\frac{n_1}{n_2}u)^{\frac{n_1+n_2}{2}}} & u > 0 \\ 0 \quad ailleurs. \end{cases}$

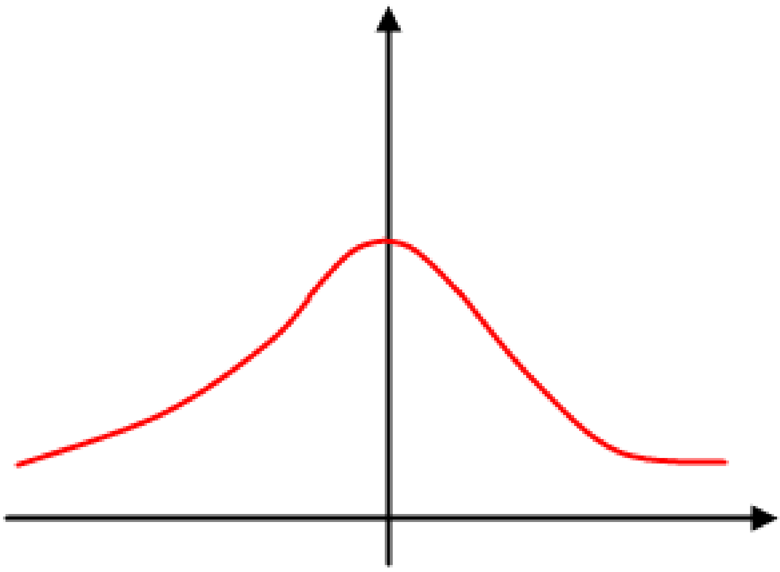
1.5 Résumé du chapitre 1

Nom de la loi	**Symbole**	**Loi**	**Ordre des valeurs**
Normale	X_i	$N(0,1)$	$\bar{X}_N$
Chi 2	χ^2_n	$\sum_i^n X_i^2$	De l'ordre du degré de liberté.
Student	S_n ou T_n	$\dfrac{N(0,1)}{\sqrt{\frac{\chi^2_n}{n}}}$	Autour de 0.
Fischer-Snedecor	F_{n_1,n_2}	$\dfrac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$	En général lorsque $u = 1$.

Il est conseillé de connaître ce tableau et de savoir le manipuler.

2 Chapitre 2 - Problèmes paramétriques

En probabilités, on utilise un modèle afin de savoir générer des nombres qui se distribuent selon la loi considérée, par exemple :

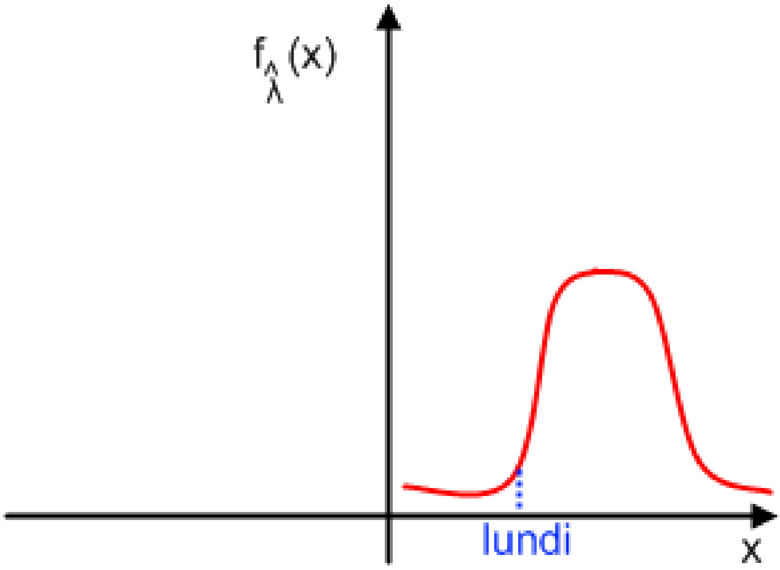


En statistiques, on fera le chemin inverse, on aura une série de mesures et à partir de celles-ci on essayera d'appliquer une loi. Par exemple :

Jour	**Voiture**	**Mesures**
Lundi	Voiture 1	x_1 secondes
" "	Voiture 2	x_2 secondes
" "	...	...
" "	Voiture 213	x_{213} secondes
Mardi	Voiture 1	x'_1 secondes
" "	Voiture 2	x'_2 secondes
" "	...	...
" "	Voiture 217	x'_{217} secondes

2.1 Méthode du maximum de vraisemblance.

A partir des mesures, on peut définir la loi de probabilité comme une fonction : $f(x, param\grave{}etres)$. Pour l'instant on considère que l'on connaît f mais on doit définir les paramètres, c'est ce que l'on appelle la statistique paramétrique. La question que l'on se pose donc est de savoir comment trouver les paramètres à partir des mesures effectuées.
Un exemple de paramètre : $\lambda_{lundi}(x_1, x_2, x_3, ..., x_{213}) = 62min13sec$,
$\lambda_{mardi}(x'_1, x'_2, x'_3, ..., x'_{217}) = 63min12sec \rightarrow \lambda$ a un caractère aléatoire.
On va avoir recours à un estimateur pour estimer le paramètre λ, $\hat{\lambda}$: $\hat{\lambda}(x_1, ..., x_n)$ v.a.r. associée à λ.

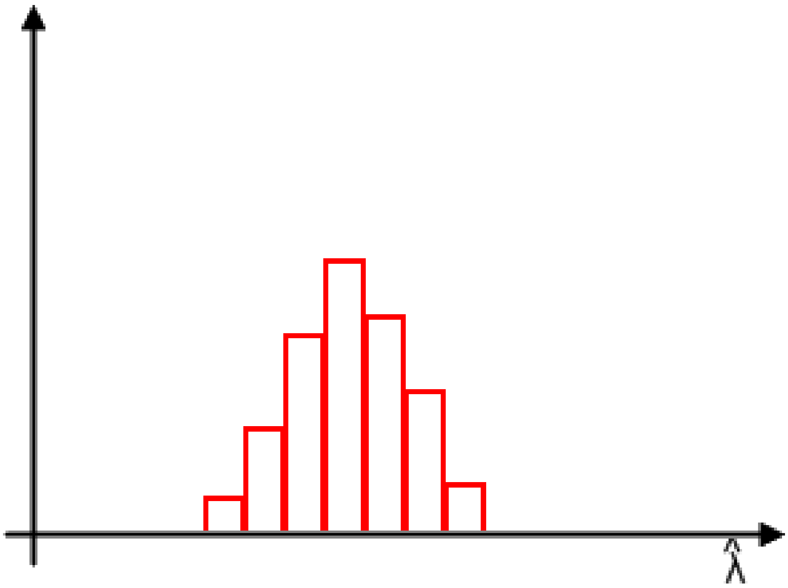


→ réalisation de l'estimateur $\hat{\lambda}(X_1, ..., X_n)$ (les X_i sont des distributions de probabilité).

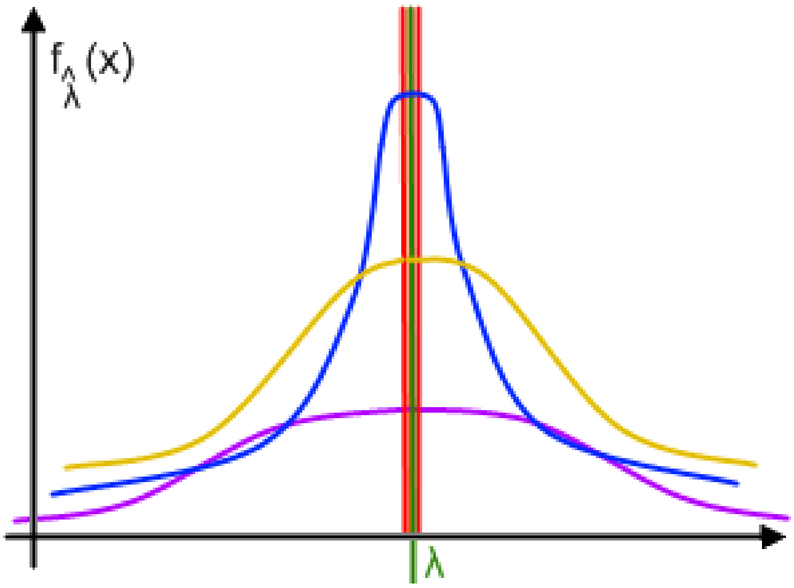
On pose v.a.r. de population X = temps d'assemblage sur la chaîne «machin» de la voiture «truc».
→ problème paramétrique : $f_\lambda(x, \lambda) \to$ nombre.
On va se donner un n-échantillons $(X_1, ..., X_n) \to \hat{\lambda}(X_1, ..., X_n)$, on définit L la fonction de vraisemblance (on suppose que chaque expérience est semblable) par : $L(x_1, ..., x_n, \lambda) = \prod_{i=1}^{n} f_{X_i}(x_i, \lambda)$

1^{er} **Jour**	$2^{\grave{e}me}$ **Jour**	...	$K^{\grave{e}me}$ **Jour**
x_1	x'_1	...	$x_1^{(k)}$
...	...	...	...
x_n	x'_m	...	$x_p^{(k)}$
$\tilde{\lambda}(x_1,...,x_n)$	$\tilde{\lambda}(x'_1,...,x'_m)$	...	$\tilde{\lambda}(x_1^{(k)},...,x_p^{(k)})$

→ histogramme des $\tilde{\lambda}$:



Propriétés :



On veut que $E(\lambda\hat{}) = \lambda$, c'est la propriété de l'estimateur **non-biaisé**.
La courbe bleue signifie que les résultats sont fort proches $\Rightarrow$ minimisation de l'erreur.
La courbe violette signifie que les résultats sont fort espacées $\Rightarrow$ maximisation de l'erreur.
La courbe orange/rouge est impossible, on va le prouver :

Inégalité de Rao-Cramer-Freschet

$\forall u$ réel, on pose $g(u) = (u(\tilde\lambda - \lambda) + \delta_\lambda \log(L))$, on a : $\forall u \in \mathbb{R}, g^2(u) \geq 0$ et on va prouver : $E\{g^2(u)\} \geq 0$.

$E\{g^2(u)\} \geq 0$
$\Leftrightarrow E\{u^2(\tilde\lambda - \lambda)^2 + 2u(\tilde\lambda - \lambda)\delta_\lambda \log(L) + (\delta_\lambda \log(L)^2\} \geq 0$
$\Leftrightarrow u^2 E\{(\tilde\lambda - \lambda)\} + 2uE\{(\tilde\lambda - \lambda)\delta_\lambda \log(L)\} + E\{(\delta_\lambda \log(L))^2\} \geq 0$
$\Leftrightarrow u^2\sigma^2(\tilde\lambda) + 2uE\{(\tilde\lambda - \lambda)\delta_\lambda \log(L)\} + E\{(\delta_\lambda \log(L))^2\} \geq 0$
Soit à simplifier : $E\{(\tilde\lambda - \lambda)\delta_\lambda \log(L)\} = E\{\tilde\lambda \frac{\delta_\lambda(L)}{L}\} - \lambda E\{\frac{\delta_\lambda(L)}{L}\}$
(dérivée d'un log par rapport à une variable = dérivée de la variable / variable)
$E\{\tilde\lambda \frac{\delta_\lambda(L)}{L}\}$
$= \int dx_1 ... \int dx_n \tilde\lambda(x_1...x_n) \frac{\delta_\lambda(L)}{L} \prod_i (f(X_i, \lambda))$
$= \delta_\lambda(\lambda)$ *(car sans biais)*
$= 1$

$\lambda E\{\frac{\delta_\lambda(L)}{L}\}$
$= \int dx_1 ... \int dx_n \frac{\delta_\lambda(L)}{L} \prod_i (f(X_i, \lambda))$
$= \delta_\lambda \int dx_1 ... \int dx_n L(x_1, ..., x_n, \lambda)$
$= \delta_\lambda(1)$
$= 0$
$\rightarrow E\{g^2(x)\} = u^2\sigma^2(\tilde\lambda + 2u + E\{(\delta_\lambda \log(L))^2\} \geq 0 \ \forall u \in \mathbb{R}$
$\rightarrow = du^2 + bu + c$

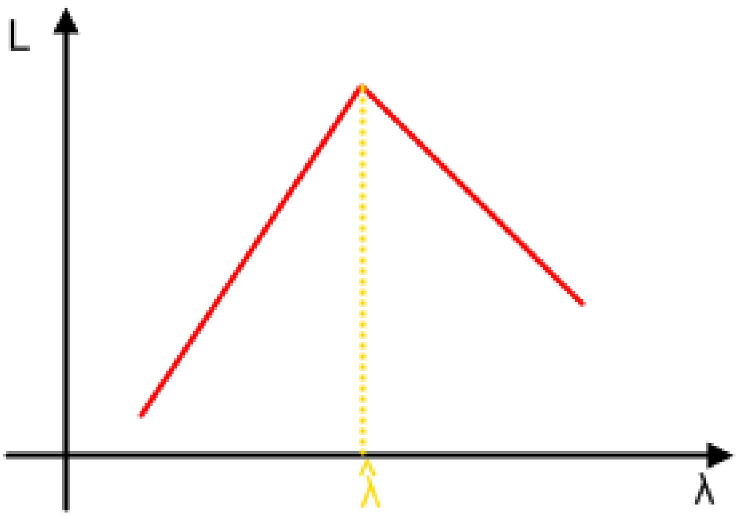
...	u_1	...	u_2	...
$+$	0	$-$	$+$	0

Il ne peut y avoir 2 racines réelles distinctes car $\geq 0 \ \forall u$, donc le discriminant $\Delta \leq 0$.
$\Delta = 4 - 4\sigma^2(\tilde\lambda)E\{(\delta_\lambda \log(L))^2\} \leq 0$
$\Leftrightarrow 1 - \sigma^2(\tilde\lambda)E\{(\delta_\lambda \log(L))^2\} \leq 0$
$\Leftrightarrow 1 \leq \sigma^2(\tilde\lambda)E\{(\delta_\lambda \log(L))^2\}$

$\Leftrightarrow 0 < \dfrac{1}{E\{(\delta_\lambda \log(L))^2\}} \leq \sigma^2(\lambda\tilde{})$
$\rightarrow$ la variance ne vaudra jamais 0 $\rightarrow \sigma > 0$, ce qui implique que la courbe rouge est impossible et que donc un estimateur sans erreur est impossible.

Construire des estimateurs

On veut un estimateur **efficace**, c'est-à-dire un estimateur :
– sans biais : $E(\tilde{\lambda}) = \lambda$
– de variance minimale : $\sigma^2(\tilde{\lambda}) = \dfrac{1}{E\{(\delta_\lambda \log(L))^2\}}$



Méthode $\tilde{\lambda}$?

$L(x_1, ..., x_n, \lambda)$ mesures réalisées, elles sont donc connues.
$\lambda\hat{}$ estimateur dont la réalisation $\tilde{\lambda}$ réalise le maximum de $L(x_1, ..., x_n, \lambda)$.

Exemple : $\hat{\lambda}$ = chaîne d'assemblage de voitures
Population λ : $f_X(x) = \dfrac{e^{(-x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \rightarrow$ on dit que $\sigma^2 = \sigma_0^2$ (on le fixe),
$\rightarrow f_X(x) = \dfrac{e^{(-x-\mu)^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}}$
$\hat{\mu}$? $\rightarrow \hat{\mu}(x_1, ..., x_n)$
$\rightarrow \prod_1^n f(x_i, \mu)$
$= \prod_1^n \dfrac{e^{(-x_i-\mu)^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}}$
$= \dfrac{e^{-\sum_{i=1}^n (x_i-\mu)^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}}$
μ ? maximum $L \rightarrow \mu$? maximum $\log(L)$ (car log est une fonction monotone croissante)
$\log(L) = -\sum_{i=1}^n (x_i - \mu)^2/2\sigma_0^2 - \dfrac{n}{2}\log(2\pi\sigma_0^2)$
$\rightarrow \delta_N \log(L)|_{\mu=\tilde{\mu}_n} = 0$ *(Equation de vraisemblance).*
$\Leftrightarrow \sum_i^n \dfrac{2(x_i - \tilde{\mu})}{2\sigma_0^2} - 0 = 0$
$\Leftrightarrow \sum_{i=1}^n (x_i - \tilde{\mu}) = 0$
$\Leftrightarrow \sum_{i=1}^n (x_i) - n\tilde{\mu} = 0$
$\Leftrightarrow \tilde{\mu} = \dfrac{1}{n}\sum_{i=1}^n (x_i)$

$\Leftrightarrow \tilde{\mu}(x_1, ..., x_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i) = \bar{x}_n$

On peut faire pareil en fixant $\mu = \mu_0$ et en cherchant σ^2, on trouve : $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2$

Imaginons que l'on ne connait rien, on a $f_X(x, \mu, \sigma^2) = \frac{e^{(-x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$

$L = \frac{e^{-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}}$

$\Leftrightarrow \log(L) = -\sum_{i=1}^{n}\left((x_i - \mu)^2/2\sigma^2\right)$

On aura max $\log(L)$ uniquement lorsque :

$$\begin{cases} \delta_{\tilde{\mu}} \log(L)|_{\tilde{\mu},\hat{\sigma}^2} = 0 & (1) \\ \delta_{\tilde{\sigma}^2} \log(L)|_{\tilde{\mu},\hat{\sigma}^2} = 0 & (2) \end{cases}$$

$(1) = +2\sum_{i=1}^{n}\frac{(x_i - \tilde{\mu})}{2\tilde{\sigma}^2}$

$\rightarrow \sum_{i=1}^{n}(x_i - \tilde{\mu}) = 0$

$\Leftrightarrow \sum_{i=1}^{n}(x_i) - n\tilde{\mu} = 0$

$\Leftrightarrow \tilde{\mu} = \frac{1}{n}\sum_{i=1}^{n}(x_i) = \bar{x}_n$

$\Leftrightarrow \boxed{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \bar{X}_n}$

$(2) = +\sum_{i=1}^{n}\frac{(x_i - \tilde{\mu})^2}{2(\tilde{\sigma}^2)^2} - \frac{n}{2\tilde{\sigma}^2}$

$\rightarrow \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{\mu})^2$

$\Leftrightarrow \boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$

Motivations *(pour utiliser cette méthode)*

1. Si, dans un problème donné $f_X(x, \lambda)$, il existe un estimateur efficace ; **alors** il est solution de l'équation de vraisemblance.
2. Si, dans un problème donné $f_X(x, \lambda)$, il **n**'existe **pas** un estimateur efficace ; **alors** la méthode du maximum de vraisemblance en fournit un qui devient asymptotiquement ($n \rightarrow \infty$) efficace.

$$\rightarrow \frac{\hat{\lambda} - \lambda}{\sqrt{var\ .\ min}} = \frac{\hat{\lambda} - \lambda}{\sqrt{E\{(\delta_\lambda \log(L))^2\}}} \rightarrow_{n\rightarrow\infty} N(0, 1)$$

On a donc, dans l'exemple plus haut, calculer les 2 paramètres de $f(x, \lambda_1, \lambda_2)$ constituant le modèle d'une série de données que l'on nous a donné (avec f connue).

On avait donc comme données de départ : x ainsi que $f(x, \lambda_1, \lambda_2) = \frac{e^{\frac{-(x-\lambda_1)^2}{2\lambda_2^2}}}{\sqrt{2\pi\lambda_2^2}}$.

On a ensuite montré que $\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ & $\hat{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

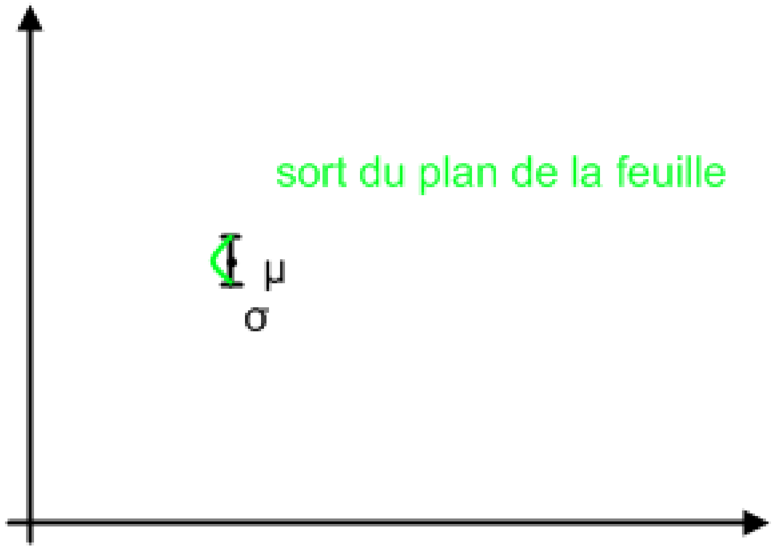
Autre exemple : Temps d'attente entre 2 coups de téléphone. *(Loi de Poisson)*

$P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!} \rightarrow$ paramètre λ.

n-échantillons : $(x_1\ ...\ x_n) \rightarrow \hat{\lambda}\ ? \rightarrow \frac{\delta}{\delta\lambda}\log(L)|_{\lambda=\hat{\lambda}} = 0$.

2.2 Méthode d'estimation par intervalle

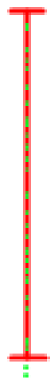
$$N : N(0, 1) \rightarrow P\{-a < N < +a\} = \int_{-a}^{+a} f_N(x)dx = \int_{-a}^{+a}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx = 2\int_0^a \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx$$



Exemple :

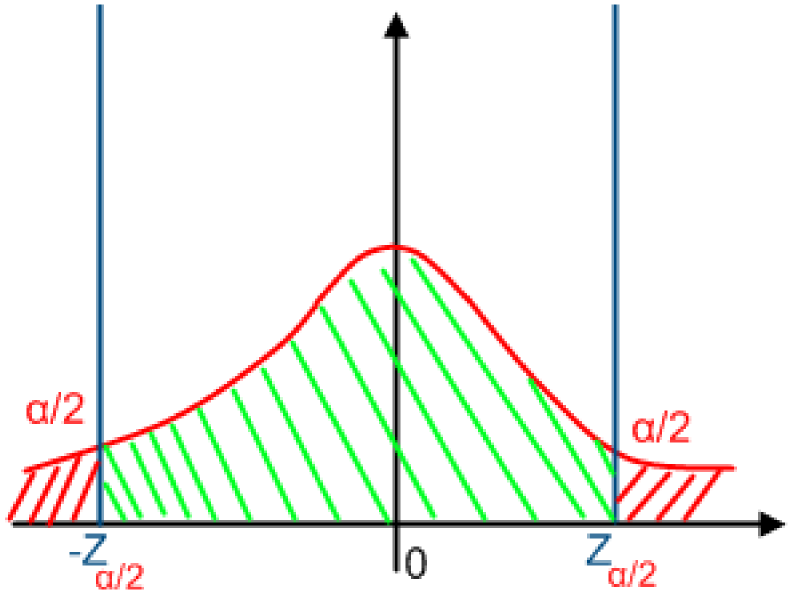
– 1 déviation standard : $P\{-1 < \frac{X-\mu}{\sigma} < +1\} = 2P\{0 < N < 1\} = 2 * 0,3413 = 0,6826 \rightarrow$ pratiquement $\frac{2}{3}$ chances qu'on tombe dans la barre d'erreur, $\frac{1}{3}$ des points seront en dehors.

– 3 déviations standards : $P\{-3 < \frac{X-\mu}{\sigma} < +3\} = 2 * 0,4987 = 0,998 \rightarrow \frac{2}{1000}$ points en dehors.



X_1 : Gaussienne : $N(\mu, \sigma) \rightarrow \bar{X}_n = \frac{1}{n}\sum_i X_i = N(\mu, \sigma/\sqrt{n})$

$\Rightarrow \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} : N(0,1)$

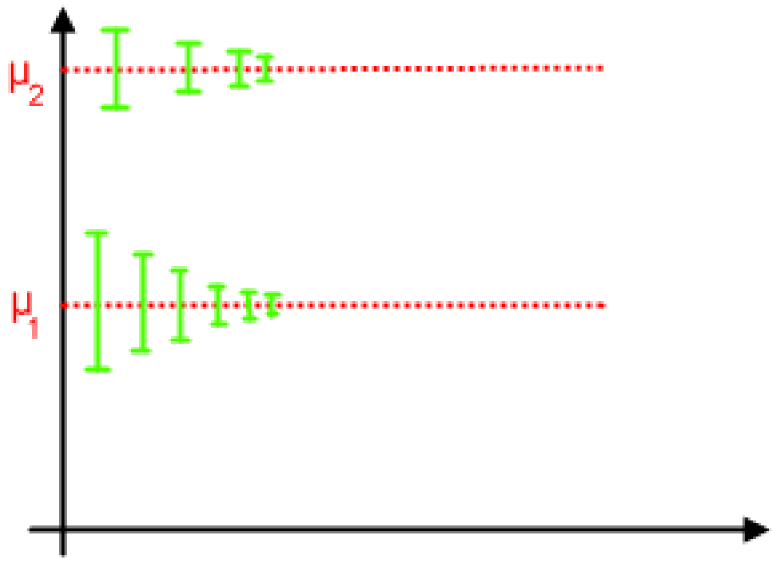


$P\{-Z_{\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < Z_{\frac{\alpha}{2}}\} = 1 - \alpha$

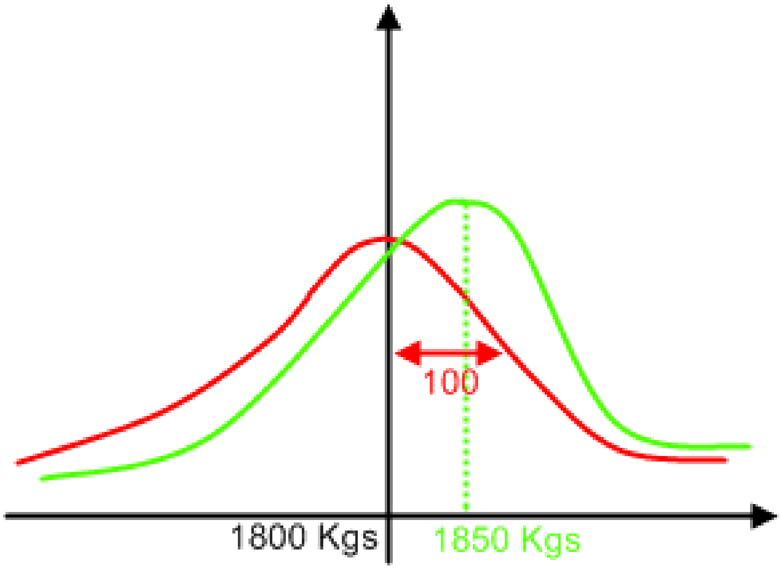
σ connu :

$P\{-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X}_n - \mu < Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\} = 1 - \alpha$

$P\{\bar{X}_n - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\} = 1 - \alpha$

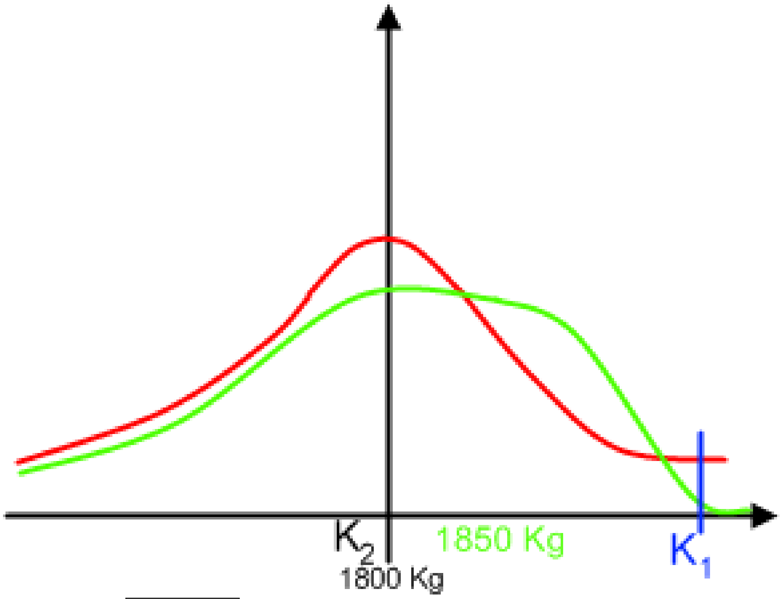


Estimation par intervalle → on veut μ avec une précision donnée.



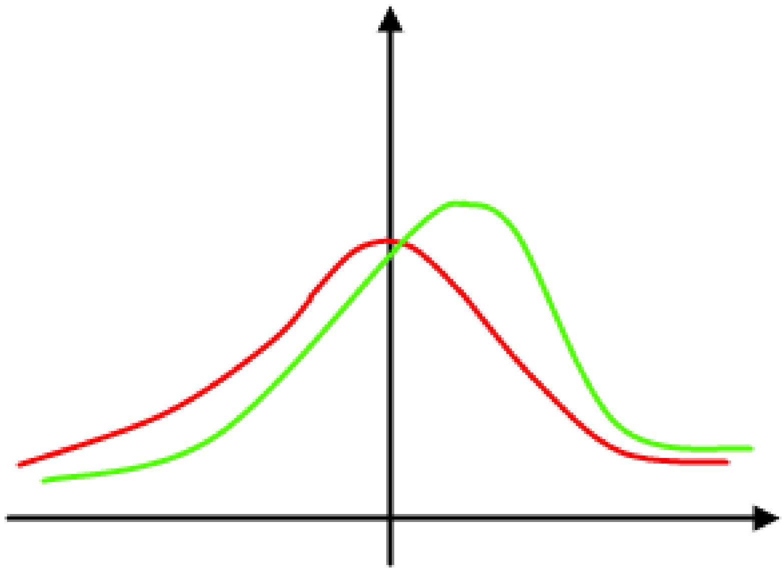
2.2.1 Tests d'hypothèse

Exemple : *Charge de rupture : 1800 kgs*, on utilise un nouveau procédé $\rightarrow$ meilleurs cables $\rightarrow$ échantillons de 50 nouveaux câbles.

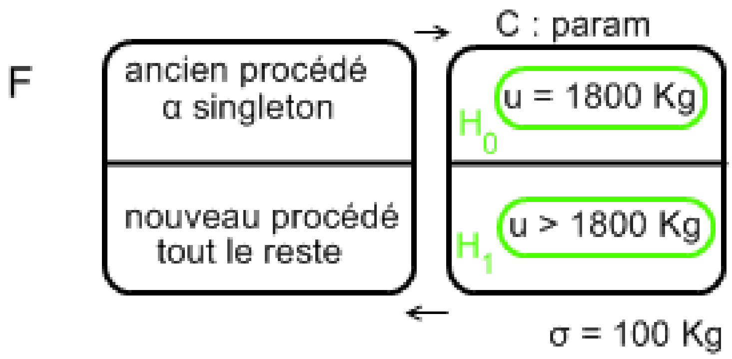


$(x_1, ..., x_n)^2 \rightarrow \boxed{\bar{X}_n > K}$? ($\rightarrow$ nouveau procédé)
$\bar{X}_n \leq K$ ($\rightarrow$ ancien procédé)

K_1	ancien procédé	nouveau procédé	K_2
accepte sur base du test	rejet de l'ancien procédé (1)	(2)	(3)
rejette sur base du test	.	(4)	(5)

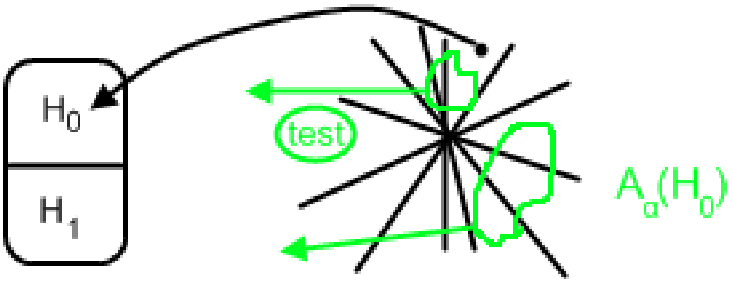


$(x_1, ..., x_{50}) \rightarrow$ modèle ?
$H_0 = H_1^c$, $H_1 = H_0^c$



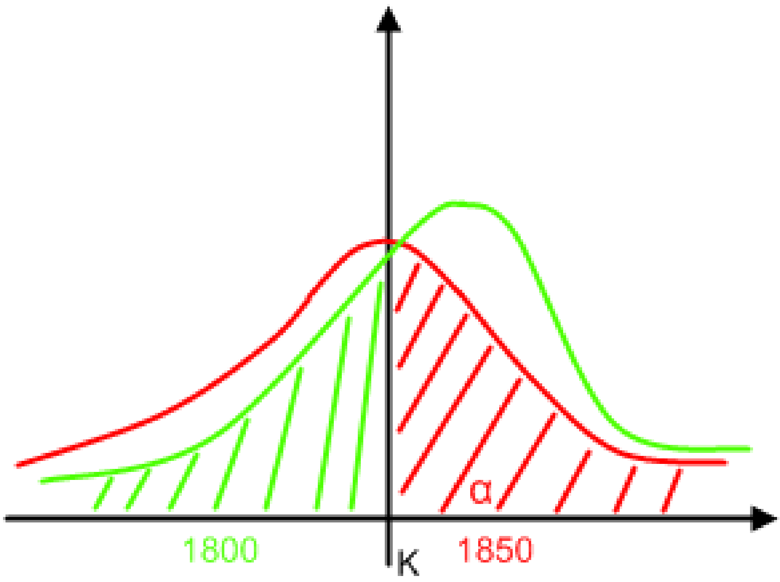
$H_0 \rightarrow AH_0$
$\quad \rightarrow RH_0 = P\{\text{Rejet } H_0 \text{ alors que } H_0 \text{ est vrai}\}$
$H_1 \rightarrow AH_1$
$\quad \rightarrow RH_1 = P\{\text{Rejet } H_1 \text{ alors que } H_1 \text{ est vrai}\}$

$\rightarrow$	H_0	H_1
RH_0	α : erreur de 1^{ere} espèce	V
RH_1	V	β : erreur de 2^{eme} espèce



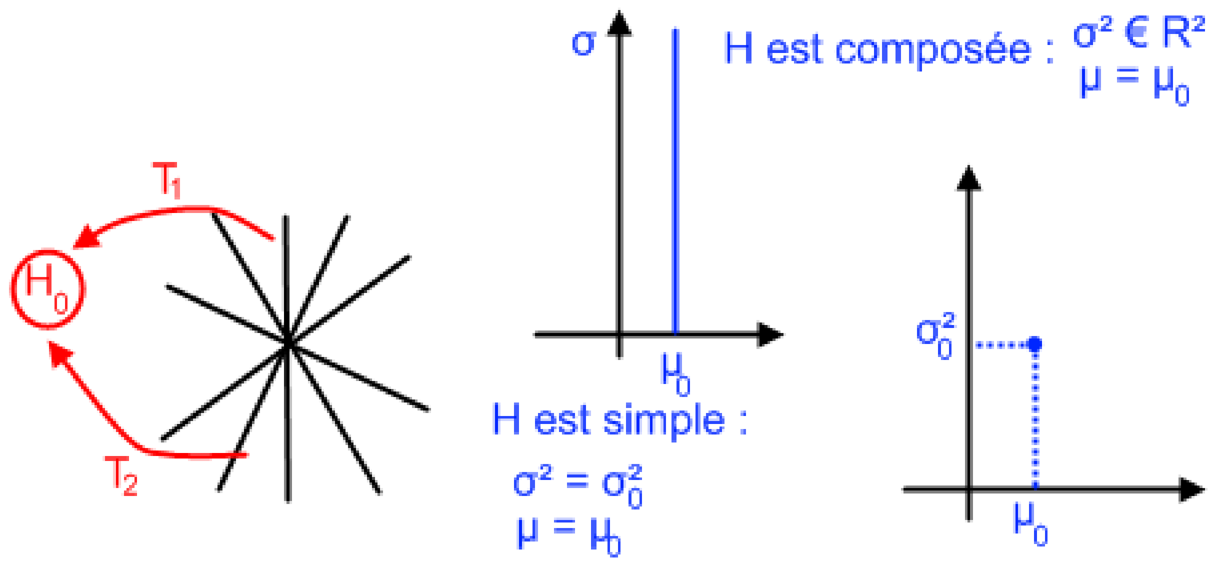
$P\{RH_0/H_0\} = \alpha$, $P\{RH_1/H_1\} = \beta$

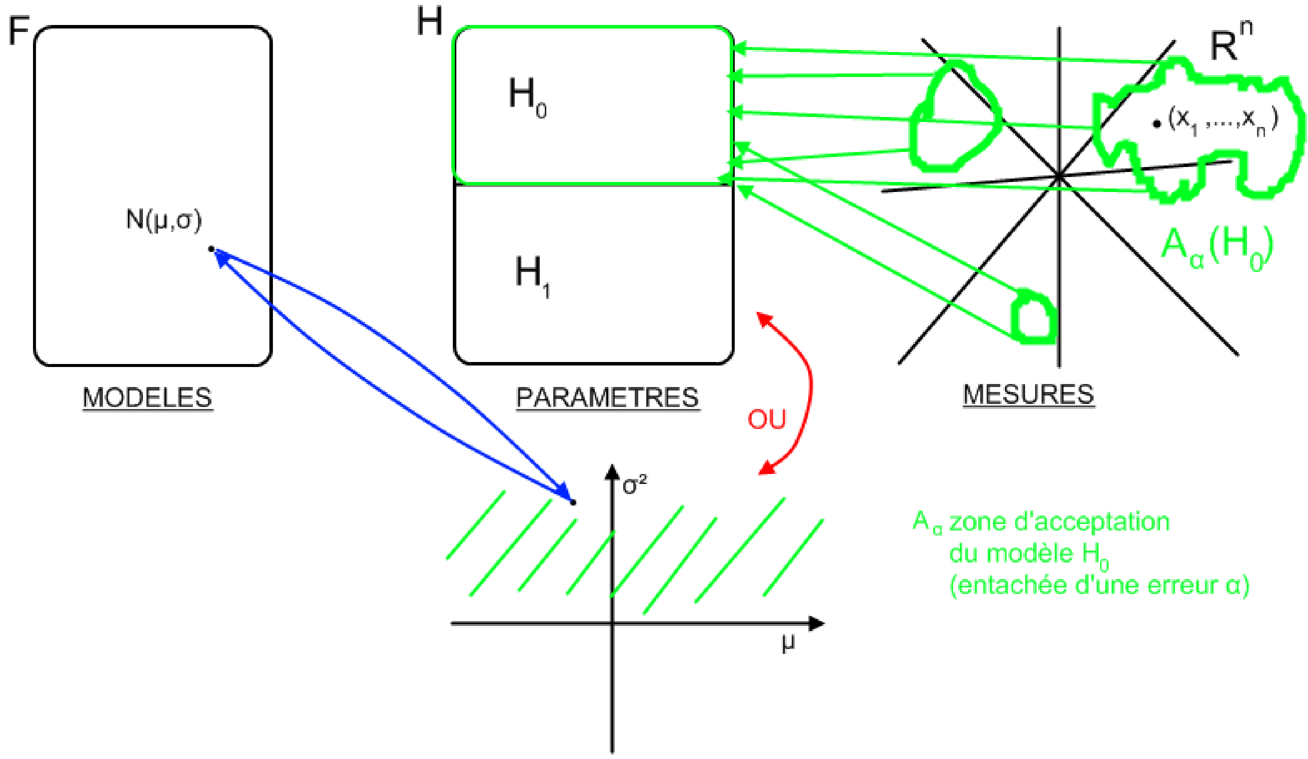
H_0, α & β, T_1 est plus puissant que T_2 ssi $\alpha_1 \leq \alpha_2$ & $\beta_1 \leq \beta_2$



Cas	H_0	H_1	Méthode
1)	simple	simple	Neyman-Pearson
2)	simple	composé	méthode du quotient de vraisemblance
3)	composé	simple	méthode du quotient de vraisemblance
4)	composé	composé	méthode du quotient de vraisemblance

H simple : $\{singleton\} \subset H$ (espace des paramètres), F : Gaussienne(μ, σ^2)
H composé : le contraire.





2.2.2 Théorème de Neyman-Pearson

On prend une population $f(x,\theta)$ avec $\theta \in \{\theta_1, \theta_2\}$ *(espace des paramètres)*.
On se donne un n-échantillons $(X_1, ..., X_n)$ qui réalise des mesures $(x_1, ..., x_n)$.
On construit la fonction de vraisemblance : $L(x_1, ..., x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$.

Hypothèses : $H_0 : \theta = \theta_1$; $H_1 : \theta = \theta_2$

Le théorème dit alors :

1. $\forall \alpha$(risque de première espèce) $\in [0,1] \exists C_\alpha \geq 0$ tel que $A_\alpha = \{(x_i...x_n) \in \mathbb{R}^n | \frac{L(x_1, ..., x_n, \theta_1}{L(x_1, ..., x_n, \theta_2} \leq C_\alpha\}$ vérifie $P_{H_0}(A_\alpha) = 1 - \alpha$
2. Le test admettant A_α comme zone d'acceptation est le plus puissant (minimise l'erreur de 2ème espèce β).

Exemple : Cartes réseau.
On a le choix entre 2 type de carte : $\mu_0 = 10$Gb/s; $\mu_1 = 20$Gb/s; on fixe $sigma_0^2 = \sigma_0^2 = 1$Mb/s. On connait la fonction de répartition : $f_X(x) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$.

$L(x_1, ..., x_n, \mu, \sigma^2) = \prod_{i=1}^{n} f_{X_i}(x_i)$

$= \prod_{i=1}^{n} \frac{e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$

$= \frac{e^{-\sum_{i=1}^{n} \frac{-(x_i-\mu)^2}{2\sigma^2}}}{(\sqrt{2\pi\sigma^2})^n}$

$$\frac{L(x_1,...,x_n,\mu_0,\sigma_0^2)}{L(x_1,...,x_n,\mu\sigma_0^2)}$$
$$= \frac{\dfrac{e^{-\sum_{i=1}^{n}\frac{-(x_i-\mu_0)^2}{2\sigma^2}}}{(\sqrt{2\pi\sigma^2})^n}}{\dfrac{e^{-\sum_{i=1}^{n}\frac{-(x_i-\mu_1)^2}{2\sigma^2}}}{(\sqrt{2\pi\sigma^2})^n}}$$
$$= \frac{e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{2\sigma_0^2}}}{e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{2\sigma_0^2}}}$$
$$= e^{-\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{2\sigma_0^2}} e^{+\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{2\sigma_0^2}} \quad (\frac{1}{e^{-a}} = e^{+a})$$
$$= e^{-\frac{1}{2\sigma_0^2}\left\{\sum_{i=1}^{n}(x_i-\mu_0)^2-\sum_{i=1}^{n}(x_i-\mu_1)^2\right\}}$$
$$= e^{-\frac{1}{2\sigma_0^2}\left\{\sum_{i=1}^{n}(x_i^2)-2\mu_0\sum_{i=1}^{n}(x_i)+\sum_{i=1}^{n}(\mu_0^2)-\sum_{i=1}^{n}(x_i^2)+2\mu_1\sum_{i=1}^{n}(x_i)-\sum_{i=1}^{n}(\mu_1^2)\right\}}$$
$$= e^{-\frac{1}{2\sigma_0^2}\left\{2(\mu_1-\mu_0)\sum_{i=1}^{n}(x_i)+n(\mu_0^2-\mu_1^2)\right\}}$$

Par Neyman-Pearson, $\exists C_\alpha$ tq $P_{H_0}\{e^{\cdots} \leq C_\alpha\} = 1-\alpha$
(On fixe le fait qu'on accepte un risque de $\alpha = 10\%$, dès lors si on examine l'égalité, on connait tout sauf C_α.)

Rappel : *Chapitre 1*
X_1 : gaussienne : $N(\mu_0,\sigma_0) \rightarrow \sum_{i=1}^{n}(X_i)/n = \bar{X}_n$
$\bar{X}_n : N(\mu_0, \frac{\sigma_0}{n}) \Rightarrow \dfrac{\bar{X}_n-\mu_0}{\frac{\sigma_0}{n}} : N(0,1)$

$\Rightarrow P_{H_0}\left\{\dfrac{-1}{2\sigma_0^2}... \leq \log C_\alpha (= C'_\alpha)\right\} = 1-\alpha$
$\Leftrightarrow P_{H_0}\left\{(x_1,...,x_n) \in \mathbb{R}^N | -\dfrac{2}{2\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^{n}(x_i) \leq C''_\alpha\right\} = 1-\alpha$ (but : se ramener à une loi du premier chapitre)
$\Leftrightarrow P_{H_0}\left\{n\dfrac{\mu_0-\mu_1}{\sigma_0^2}\dfrac{1}{n}\sum_{i=1}^{n} x_i \leq C''_\alpha\right\} = 1-\alpha$ (on multiplie par n et on divise par n)
$\Leftrightarrow P_{H_0}\left\{\dfrac{\mu_0-\mu_1}{\frac{\sigma_0^2}{n}}\bar{X}_n \leq C''_\alpha\right\} = 1-\alpha$

1. $\underline{\mu_0 > \mu_1} \rightarrow P_{H_0}\left\{\dfrac{\bar{X}_n-\mu_0}{\sigma_0/\sqrt{n}} \leq C'''_\alpha\right\} = 1-\alpha$ (ici, par les tables, $C'''_\alpha = 1,28$)
2. $\underline{\mu_1 > \mu_0} \rightarrow P_{H_0}\left\{\dfrac{\bar{X}_n-\mu_0}{\sigma_0/\sqrt{n}} \geq C'''_\alpha\right\} = 1-\alpha$ (ici, par les tables, $C'''_\alpha = 1,28$)

2.2.3 Méthode du quotient de vraisemblance

On introduit $\Lambda_n = \dfrac{\sup_H L}{\sup_{H_0} L}$ et on dit qu'on est :
- dans la zone **d'acceptation** si $P\{\Lambda_n \leq \lambda_\alpha\} = 1-\alpha$
- dans la zone **de rejet** si $P\{\Lambda_n > \lambda_\alpha\} = \alpha$

Exemple : (voir syllabus p70)
On prend une population Gaussienne, et on a 2 hypothèses : $H_0 : \mu = \mu_0$ et $H_1 : \mu \neq \mu_O$.
On construit la fonction de vraisemblance : $\prod_{i=1}^{n}\dfrac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} = \dfrac{e^{-\sum_{i=1}^{n}\frac{-(x_i-\mu)^2}{2\sigma^2}}}{(\sqrt{2\pi\sigma^2})^n}$.

numérateur : $\sup_H L = sup_{(\mu,\sigma^2)} L = L(x_1...x_n, \hat{\mu}, \hat{\sigma}^2)$

avec $\hat{\mu} = \dfrac{1}{n}\sum_{i=1}^n x_i$ et $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2$

dénominateur : $\sup_H L = sup_{(\mu=\mu_0,\sigma^2)} L = L(x_1...x_n, \mu_0, \hat{\sigma}_2^2)$

avec $\hat{\sigma}_2^2 = \dfrac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2$

$\Rightarrow \Lambda_n = \frac{numerateur}{denominateur}$

$$= \frac{L(x_1...x_n, \hat{\mu}, \hat{\sigma}^2)}{L(x_1...x_n, \mu_0, \hat{\sigma}_2^2)} = \frac{\dfrac{e^{-\sum_{i=1}^n \frac{(x_i-\bar{x}_n)^2}{2\left(\frac{1}{n}\sum_{i=1}^n (x_i-\bar{x}_n)^2\right)}}}{\left(\sqrt{\frac{2\pi}{n}\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right)^n}}{\dfrac{e^{-\sum_{i=1}^n \frac{(x_i-\mu_0)^2}{2\left(\frac{1}{n}\sum_{i=1}^n (x_i-\mu_0)^2\right)}}}{\left(\sqrt{\frac{2\pi}{n}\sum_{i=1}^n (x_i-\mu_0)^2}\right)^n}}$$

Les $e^{-\cdots}$ se neutralisent car se simplifient toutes les 2 en $e^{n/2}$

$$= \left(\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right)^{n/2}$$

A faire donc : $P\left\{A_\alpha | \left(\dfrac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right)^{n/2} \leq \lambda_\alpha\right\} = 1-\alpha$

$\Leftrightarrow P\left\{A_\alpha | \left(\dfrac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right) \leq (\lambda_\alpha)^{2/n} = \lambda'_\alpha\right\} = 1-\alpha$

$\Rightarrow$ Se ramener à une loi du chapitre 1.

On cherche à simplifier :

$$\left(\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right)$$

$$= \left(\frac{\sum_{i=1}^n (x_i-\bar{x}_n+\bar{x}_n-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right)$$

$$= \frac{\sum_{i=1}^n (x_i-\bar{x}_n)^2 + 2(\bar{x}_n-\mu_0)\sum_{i=1}^n (x_i-\bar{x}_n) + n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}$$

$$= \frac{\sum_{i=1}^n (x_i-\bar{x}_n)^2 + \left[2(\bar{x}_n-\mu_0)\left(\sum_{i=1}^n (x_i) - \sum_{i=1}^n \bar{x}_n\right)\right] + n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}$$

$$= \frac{\sum_{i=1}^n (x_i-\bar{x}_n)^2 + \left[2(\bar{x}_n-\mu_0)\left(\sum_{i=1}^n (x_i) - n\bar{x}_n\right)\right] + n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}$$

$$= \frac{\sum_{i=1}^n (x_i-\bar{x}_n)^2 + \left[2(\bar{x}_n-\mu_0)\left(\sum_{i=1}^n (x_i) - n\frac{1}{n}\sum_{i=1}^n (x_i)\right)\right] + n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2} \text{ (par définition de } \bar{x}_n)$$

$$= \frac{\sum_{i=1}^n (x_i-\bar{x}_n)^2 + \mathbf{0} + n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}$$

$$= 1 + \frac{n(\bar{x}_n - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$$

$$\Rightarrow P\left\{A_\alpha | \frac{n(\bar{x}_n - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} \leq \lambda'_\alpha - 1 = \lambda''_\alpha\right\} = 1 - \alpha$$

$$\Leftrightarrow P\left\{-\lambda'''_\alpha \leq \frac{\bar{x}_n - \mu_0}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}} \leq \lambda'''_\alpha\right\} = 1 - \alpha$$

$$\bar{X}_n : N(\mu, \sigma_0/\sqrt{n}) \Rightarrow \frac{\bar{X}_n - \mu_0}{\sigma_0/\sqrt{n}} : N(0,1)$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{X_i - \mu_0}{\sigma_0}\right)^2 = \chi^2_n$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}_n}{\sigma_0}\right)^2 = \chi^2_{n-1}$$

$$P\left\{-\lambda''''_\alpha \leq \frac{\sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma_0}}{\sqrt{\frac{\sum_{i=1}^{n}\frac{(X_i - \bar{X}_n)^2}{\sigma_0}}{n-1}}} \leq \lambda''''_\alpha\right\} = \alpha - 1$$

$$\Rightarrow P\left\{-\lambda''''_\alpha \leq \frac{\sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma_0}}{\sqrt{\chi^2_{n-1}/n-1}} \leq \lambda''''_\alpha\right\} = \alpha - 1$$

$$\Rightarrow P\left\{-\lambda''''_\alpha \leq \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/n-1}} \leq \lambda''''_\alpha\right\} = \alpha - 1$$

$$\Rightarrow P\left\{-\lambda''''_\alpha \leq S_{n-1} \leq \lambda''''_\alpha\right\} = \alpha - 1$$

On a dès lors plus qu'à avoir recours aux tables et ainsi finir l'exercice.

3 Chapitre 3

4 Chapitre 4