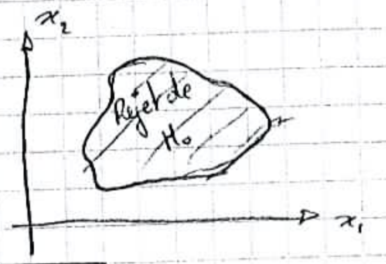
Chap 2 : Tests statistiques

I Généralités :

Un test statistique est un mécanisme qui permet de décider à partir d'un échantillon $X = (X_1 \dots X_n)$ entre deux (ou plus) hypothèses notées H_0 et H_1. (ou $H_0, H_1, H_2 \dots$)

H_0 s'appelle l'hypothèse nulle. Tandis que H_1 s'appelle l'hypothèse alternative.



Rq : On se limite dans ce cours à 2 hypothèses

> Elaborer une stratégie de Test, c'est déterminer une statistique notée $T(X_1, \dots, X_n)$ telle que :
>
> si $T(x_1 \dots x_n) \in \Delta$ alors on rejette l'hypothèse H_0 (c'est à dire on accepte H_1)
>
> Si $T(x_1 \dots x_n) \notin \Delta$ alors on accepte l'hypothèse H_0

Vocabulaire : $\{(x_1 \dots x_n) \,/\, \underbrace{T(x_1 \dots x_n) \in \Delta}_{\text{on rejette } H_0}\}$ s'appelle la région critique du test

On distinguera dans ce cours les tests dits paramétriques des tests non paramétriques. Un test est paramétrique si la forme de la loi des VA X_i est connue et que l'on cherche à tester la valeur d'un ou de plusieurs de ses paramètres.

Exemple : $X_i \sim N(m, \sigma^2)$ σ^2 connue

Test n°1 H_0 $m = 1$ H_1 $m = 2$

On parle d'hypothèses simples lorsqu'elles sont réduites à un point. (1 seule valeur du paramètre)

le test n°1 est un test d'hypothèses simples.

Test n°2 : (H₀) $m > 0$ (H₁) $m \leq 0$

Ce test est un test paramétique à hypothèses appelées hypothèses Composées.

Test n°3 : (H₀) $X \sim N(m, \sigma^2)$ (H₁) non H₀

Ce test est non paramétique car il porte sur la loi des va X_i et non pas sur ses paramètres.

Rq : Pour le test n°1 on pourrait envisager la stratégie suivante

Rejet de H₀ si $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i > 1,5$

La région critique est séparée par un hyperplan (droite pour $n = 2$, plan pour $n = 3$.

Pour déterminer la performance d'un test, on s'interessera aux risques de se tromper :

① $\boxed{\text{risque de } 1^{ère} \text{ espèce} \quad \text{risque } \alpha}$ $\alpha = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$

② $\boxed{\text{risque de } 2^{nde} \text{ espèce} \quad \text{risque } \beta}$ $\beta = P[\text{Rejeter } H_1 / H_1 \text{ vraie}]$

Un test sera d'autant meilleur que α et β seront "petits"

α et β jouent ~~des~~ rôles parfaitement symétriques et $\boxed{\text{pourtant}}$:

$\boxed{\text{Ex 1 :}}$ (H₀) le patient est sain

(H₁) le patient est atteint d'une maladie X

(Rq : le patient est sain signifie absence de maladie $\Rightarrow$ hypothèse nulle)

$\alpha = P[$ on décide que le patient est malade / le patient est sain $]$
probabilité de fausse alarme (PFA)

$\beta = P[$ on décide que le patient est sain / le patient est malade $]$
probabilité de non détection (PND)

Dans la mesure où on a le choix, on privilégiera β

Ex 2 : H_0 non H_1 (absence d'avion)
H_1 un avion bombarde l'N7 (présence d'avion)

α = PFA
β = PND

Définition : On appelle puissance du test $\Pi = 1 - \beta$
(probabilité de détection)

II Exemple :

$X_i \sim N(m, \sigma^2)$ σ^2 connue
$(X_1, \dots, X_n)$ échantillon H_0 $m = m_0$ H_1 $m = m_1 > m_0$

Etudions la stratégie suivante :

$$\boxed{\text{Rejet de } H_0 \text{ si } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i > S_\alpha}$$

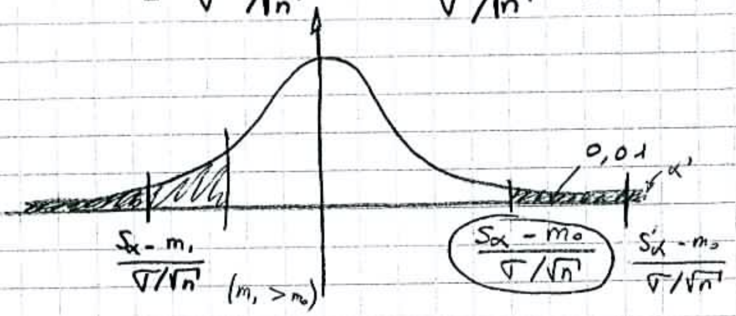
(seuil S_α)
dépend du risque α

Détermination de S_α : On se fixe α (en général 0,1 ; 0,5 ; 0,01 ; ou 0,05,

$\alpha = 0{,}01 = P[$ rejeter H_0 / H_0 vraie $]$
$= P[\overline{X} > S_\alpha / m = m_0]$

On sait que $\overline{X} \sim N\left(m, \frac{\sigma^2}{n}\right)$
$= P\left[\overline{X} > S_\alpha / \overline{X} \sim N\left(m_0, \frac{\sigma^2}{n}\right)\right]$

$$\alpha = 0{,}01 = P\left[U = \frac{\overline{X} - m_0}{\sigma/\sqrt{n}} > \frac{S_\alpha - m_0}{\sigma/\sqrt{n}} \;/\; U \sim N(0,1)\right] = \int_{\frac{S_\alpha - m_0}{\sigma/\sqrt{n}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$



les tables donnent $\frac{S_\alpha - m_0}{\sigma/\sqrt{n}} = \dots$

d'où $S_\alpha = \dots$

$$\boxed{\text{plus } \alpha \searrow, \text{ plus } S_\alpha \nearrow}$$

<u>Calcul de β</u> (ou de $\Pi = 1 - \beta$)

$$\beta = P[\text{rejeter } H_1 / H_1 \text{ vraie}] = P[\overline{X} < S_\alpha / m = m_1]$$

$$= P\left[V = \frac{\overline{X} - m_1}{\sigma/\sqrt{n}} < \frac{S_\alpha - m_1}{\sigma/\sqrt{n}} \;/\; V \sim N(0,1)\right]$$

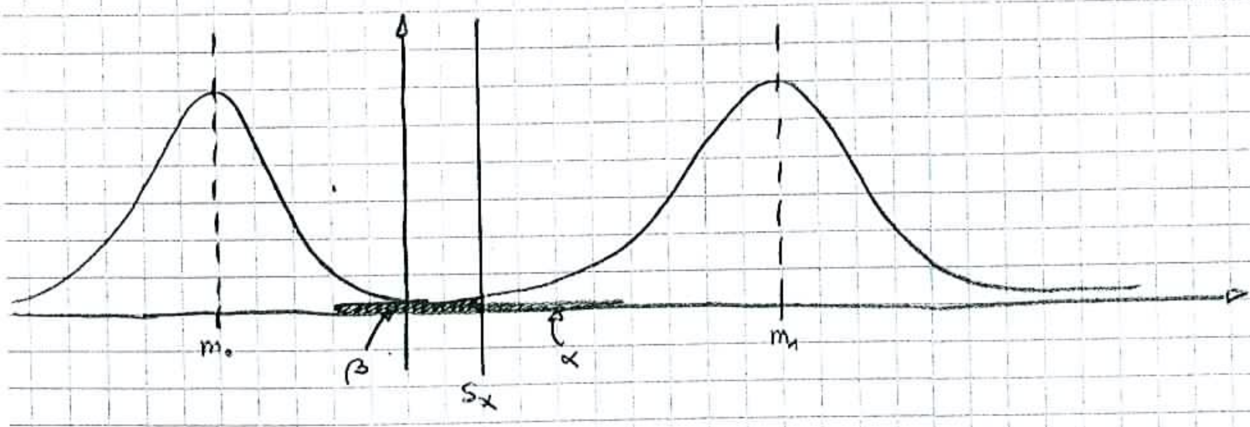
On détermine $\beta = \int_{-\infty}^{\frac{S_\alpha - m_1}{\sigma/\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$ à l'aide des tables de la loi normale.

plus $\alpha \searrow$, plus $S_\alpha \nearrow$ et donc plus $\beta \nearrow$

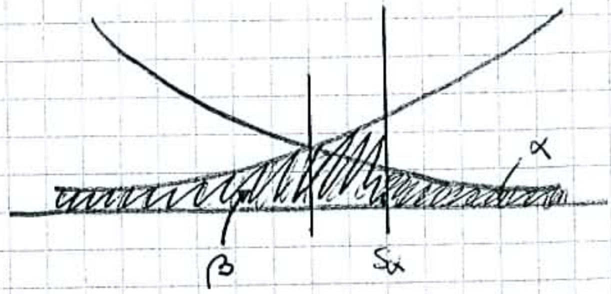
<u>α et β varient en sens inverse</u> (<u>β fct° décroissante de α</u>)

<u>Autre représentation</u>, $\alpha = P\left[\overline{X} > S_\alpha \;/\; \overline{X} \sim N\left(m_0, \frac{\sigma^2}{n}\right)\right]$

$\beta = P\left[\overline{X} < S_\alpha \;/\; \overline{X} \sim N\left(m_1, \frac{\sigma^2}{n}\right)\right]$



Zoom



Si on cherche à minimiser $\alpha + \beta$, il faut placer S_α à l'intersection des 2 courbes.

III Théorème (ou LEMME) de NEYMAN-PEARSON (1933)

(NP)

L'approche de NP consiste à fixer α et à chercher un test qui minimise β. (pour cette valeur de α fixé). On dit souvent que le test de NP est optimal.

III.1. Cas de variables aléatoires continues et d'hypothèses simples.

H_0 $\theta = \theta_0$ $\qquad$ H_1 $\theta = \theta_1$

X_i possède une densité de probabilité sous chacune des 2 hypothèses notées

$$\underbrace{f(x_i ; \theta_0)}_{\text{densité de } X_i \text{ sous } H_0} \quad \text{et} \quad \underbrace{f(x_i ; \theta_1)}_{\text{densité de } X_i \text{ sous } H_1}$$

A α fixé, le test qui minimise β est de la forme

Rejet de H_0 si $\dfrac{L(x_1, \dots, x_n / H_1)}{L(x_1, \dots, x_n / H_0)} > S_\alpha$

Rq: $L(x_1, \dots, x_n / H_i) = \prod_{j=1}^{n} f(x_j, \theta_i) = f(x_1, \dots, x_n ; \theta_i)$

Exemple: $X_i \sim N(m, \sigma^2)$ $\quad \sigma^2$ connue

H_0 $m = m_0$ $\qquad$ H_1 $m = m_1$

Le test de NP est défini par

Rejet de H_0 si $$\frac{\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x_j - m_1)^2}{2\sigma^2}}{\prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x_j - m_0)^2}{2\sigma^2}} > S_\alpha$$

si $\quad \exp -\frac{1}{2\sigma^2}\left[\sum_{j=1}^{n}(x_j - m_1)^2 - \sum_{j=1}^{n}(x_j - m_0)^2\right] > S_\alpha$

(on prend le ln)

si $\quad \sum_{j=1}^{n}\left(\cancel{x_j^2} - 2m_1 x_j + m_1^2 - \cancel{x_j^2} + 2m_0 x_j - m_0^2\right) < k\alpha$

$$\underbrace{-2(m_1 - m_0)n}_{<0}\,\bar{x} + n\left(m_1^2 - m_0^2\right) < k_\alpha$$

$\boxed{\text{Si } m_1 > m_0}$ $\quad \boxed{\text{Rejet de } H_0 \text{ si } \bar{x} > \mu_\alpha}$

$\boxed{\text{Si } m_1 < m_0}$ $\quad \boxed{\text{Rejet de } H_0 \text{ si } \bar{x} < \mu_\alpha}$

On retrouve la stratégie du § II

<u>Rq</u>: Effectuer un test de NP c'est :

1) déterminer la zone de rejet de H_0 (zone critique) région

2) On se fixe α (= 0,01 par ex) et on détermine μ_α

3) On calcule la statistique de test (ici $\bar{x} = \frac{1}{n}\sum x_i$) que l'on compare à μ_α et on accepte ou rejette l'hypothèse H_0 avec le risque α = –

4) On calcule β

17/01/2000

Une représentation habituelle permettant d'analyser les performances d'un test consiste à tracer **les courbes COR (Caractéristiques Opérationnelles du Récepteur)** définies par $PD = 1-\beta = f(\alpha)$
↳ probabilité de Détection

$\alpha = PFA$
$= P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$

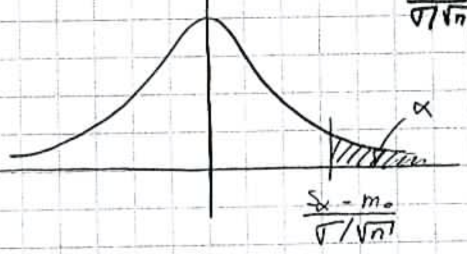
$\beta = PND$
$= P[\text{Rejeter } H_1 / H_1 \text{ vraie}]$

__exemple__ : On se fixe $\alpha = P[\overline{X} > S_\alpha / m = m_0]$

$$= P\left[\overline{X} > S_\alpha / \overline{X} \sim N\left(m_0, \frac{\sigma^2}{n}\right)\right]$$

$$= P\left[\underbrace{\frac{\overline{X} - m_0}{\sigma/\sqrt{n}}}_{U} > \frac{S_\alpha - m_0}{\sigma/\sqrt{n}} \;/\; U \sim N(0,1)\right]$$

$$= \int_{\frac{\overline{X}-m_0}{\sigma/\sqrt{n}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du$$



$$\int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du = erfc(x)$$ __Tabulé__
↳ fct° MATLAB

$+ x \nearrow$, $+ erfc \searrow$

Donc $\frac{S_\alpha - m_0}{\sigma/\sqrt{n}} = erfc^{-1}(\alpha)$

$$\boxed{S_\alpha = m_0 + \frac{\sigma}{\sqrt{n}} erfc^{-1}(\alpha)}$$

$PD = 1 - \beta = P[\text{accepter } H_1 / H_1 \text{ vraie}]$

$= P[\text{rejeter } H_0 / H_1 \text{ vraie}]$

$= P\left[\overline{X} > S_\alpha / \overline{X} \sim N\left(m_1, \frac{\sigma^2}{n}\right)\right]$

$= \boxed{erfc\left(\frac{S_\alpha - m_1}{\sigma/\sqrt{n}}\right) = PD}$

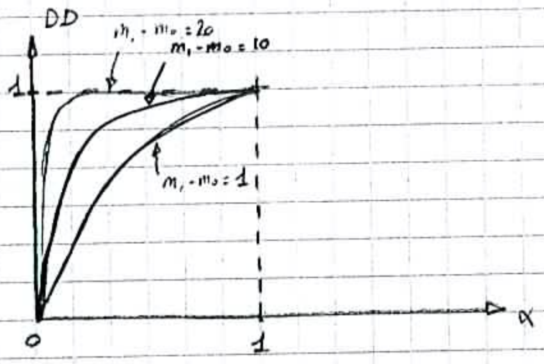
Courbes COR : $\boxed{PD = erfc\left[\frac{-\sqrt{n}\,(m_1 - m_0)}{\sigma} + erfc^{-1}(\alpha)\right]}$

__Interprétations__ : ① plus $n \nearrow$ plus $erfc \nearrow$ plus $PD \nearrow$

qd $n \to +\infty$ $PD \to 1$

(2) plus $m_1 - m_0$ ↗ (H₀ et H₁ "éloignés") plus PD ↗

(3) plus σ ↘ plus PD ↗



III . 2 Cas discret : ($X = (X_1 \dots X_n)$ est un échantillon de VA discrètes)

Lemme de Neyman - Pearson :

Parmi tous les tests dont le risque de 1ère espèce est $\leq \alpha$, il en existe un de puissance ($PD = 1-\beta$) maximale défini par

Rejet de H_0 si $\dfrac{L(x_1 \dots x_n / H_1)}{L(x_1 \dots x_n / H_0)} > S_\alpha$

Exemple : $X_i \sim \mathcal{P}(\lambda)$ $\quad$ (H_0) $\lambda = \lambda_0$ $\quad$ (H_1) $\lambda = \lambda_1$ $\quad$ ($\lambda_1 > \lambda_0$)

Rejet de H_0 si $\dfrac{P[X_1 = x_1, \dots, X_n = x_n / H_1]}{P[X_1 = x_1, \dots, X_n = x_n / H_0]} > S_\alpha$

si $\dfrac{\prod_{i=1}^{n} e^{-\lambda_1} \dfrac{\lambda_1^{x_i}}{\cancel{x_i!}}}{\prod_{i=1}^{n} e^{-\lambda_0} \dfrac{\lambda_0^{x_i}}{\cancel{x_i!}}} > S_\alpha$

$-n(\lambda_1 - \lambda_0) + \sum x_i (\ln \lambda_1 - \ln \lambda_0) > \ln S_\alpha$

$$\boxed{\text{Rejet de } H_0 \text{ si } \sum_{i=1}^{n} x_i > \mu_\alpha}$$

A.N. : $n = 2$ $\lambda_0 = 1$ $\lambda_1 = 2$ $\alpha = 0{,}05$

$\alpha = 0{,}05 = P[\text{Rejeter } H_0 / H_0 \text{ vraie}] = P[\sum_{i=1}^{2} X_i > \mu_\alpha / \lambda = \lambda_0]$

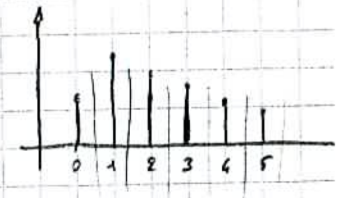
Si $\lambda = \lambda_0$ $X_i \sim P(\lambda_0)$ alors on montre que $X_1 + X_2 \sim P(2\lambda_0)$

$\alpha = 0{,}05 = P[U > \mu_\alpha / U \sim P(2)]$

Si $\mu_\alpha \leq 0$ $P[U > \mu_\alpha] = 1$ $\alpha = 1$

si $\mu_\alpha \in]0,1]$ $P[U > \mu_\alpha] = 1 - P[U = 0] = 1 - \frac{(2\lambda_0)^0}{0!} e^{-2\lambda_0} \# 0{,}865$ Table

si $\mu_\alpha \in]1,2]$ $P[U > \mu_\alpha] = 1 - P[U = 0 \text{ ou } U = 1]$ ...



...

Si $\mu_\alpha \in]3,4]$ $P[U > \mu_\alpha] = P[U > 4] = 0{,}0527$

$\boxed{\text{Si } \mu_\alpha \in]4,5] \quad P[U > \mu_\alpha] = 0{,}0166}$

Conclusion : Le test de risque $\alpha \leq 0{,}05$ le plus puissant est
Rejet de H_0 si $x_1 + x_2 > 5$
il admet un risque de 1ère espèce $\alpha \# 0{,}0166$

Rq : Si n grand, on approche la loi de U par une loi normale en vertu du théorème de la limite centrale
ici on aurait $U = \sum_{i=1}^{n} X_i \tilde{\#} \mathcal{N}(n\lambda, n\lambda)$

IV Test du rapport des vraisemblances généralisé :

(ou Test GLR (Generalized Likelihood ratio))

Ce test est adapté aux hypothèses composées $\begin{cases} H_0 & \theta \in A \\ H_1 & \theta \in B \end{cases}$

où les ensembles A et B ne sont pas tous deux réduits à un point.

Le test GLR est défini par

$$\text{Rejet de } H_0 \text{ si } \frac{L(x_1, \dots, x_n ; \widehat{\theta_{1mv}})}{L(x_1, \dots, x_n ; \widehat{\theta_{0mv}})} > S_\alpha$$

où $\hat{\theta}_{0mv}$ et $\hat{\theta}_{1mv}$ sont les estimateurs du maximum de vraisemblance de θ sous les hypothèses H_0 et H_1 respectivement.

Rq: $$\frac{L(x_1, \dots, x_n ; \widehat{\theta_{1mv}})}{L(x_1, \dots, x_n ; \widehat{\theta_{0mv}})} = \frac{\sup_{\theta \in B} L(x_1 \dots x_n ; \theta)}{\sup_{\theta \in A} L(x_1 \dots x_n ; \theta)}$$

V Test du χ^2 :

Le test du χ^2 est un test d'ajustement (ou d'adéquation) non paramétique, de la forme H_0 $L = L_0$ une certaine loi

H_1 $L \neq L_0$

On se propose de tester si l'échantillon $(X_1 \dots X_n)$ est de loi L_0 (discrète ou continue) ou non.

Principe: On découpe le support de la loi L_0 connue en k parties appelées classes et notées $C_1 \dots C_k$.

On note $P_{0k} = P[X_j \in C_k / X_j \sim L_0]$

Exemple: $X_j \sim N(m, \sigma^2)$ $m=0$ $\sigma^2 = 1$ $C_1 =]-\infty, 0[$ $C_2 = [0,1]$ $C_3 =]1, +\infty[$

$P_{01} = P[X_j \in C_1] = \frac{1}{2}$ $P_{02} = \int_0^1 \frac{1}{\sqrt{2\pi\sigma^2}} \exp() \, dx$

$P_{03} = \int_1^{+\infty} \_\_\_$

On définit les va Y_j^k de la façon suivante :

$$Y_j^k = \begin{cases} 1 & \text{si } x_j \in C_k \\ 0 & \text{si } x_j \notin C_k \end{cases}$$

Alors $Z_k = \sum_{j=1}^{n} Y_j^k$ représente la va "nbre d'observations $\in C_k$"

D'après le théorème de la limite centrale, (les va Y_j^k sont ind. dans la mesure où les x_j le sont); pour <u>n grand</u>

$$\boxed{Z_k \underset{\#}{\approx} \mathcal{N}(nP_{0k}, nP_{0k}(1-P_{0k}))}$$ car $E[Y_j^k] = 1.P[x_j \in C_k] + 0.P[x_j \notin C_k]$

Var $Y_j^k = P_{0k}(1-P_{0k})$ $\left(= E[(Y_j^k)^2] - E[Y_j^k]^2 = P_{0k} - (P_{0k})^2\right)$

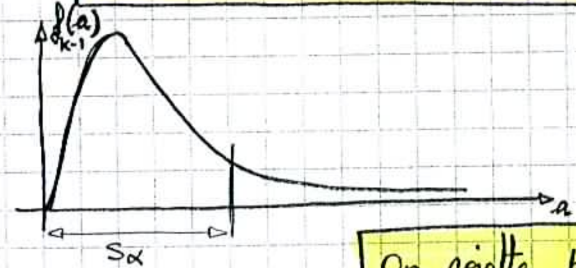
De manière équivalente $U_k = \dfrac{Z_k - nP_{0k}}{\sqrt{nP_{0k}}} \sim N(0, 1-P_{0k})$

<u>Rq</u>: les VA U_k ne sont pas indépendantes car $\sum_{k=1}^{n} Z_k = n$

(n = nbre de x_i)

On admettra que sous l'hypothèse H_0 ($x_i \sim L_0$)

$$\boxed{A = \sum_{k=1}^{k} \frac{(Z_k - nP_{0k})^2}{nP_{0k}} \xrightarrow[n \to \infty]{L} \chi^2_{k-1}}$$



$$\boxed{\text{On rejette } H_0 \text{ si } A > S_\alpha}$$

Le calcul de S_α se fait comme suit, on se fixe α (par ex 0,01)

$$\alpha = 0{,}01 = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$$
$$= P[A > S_\alpha / X_i \sim L_0] = \int_{S_\alpha}^{+\infty} f_{k-1}(a)\, da$$ → <u>Tables du χ^2</u>

les tables du χ^2 donnent $\underset{\xi}{S_\alpha}$

$$\sum_{k=1}^{K} \frac{(Z_k - np_{0k})^2}{np_{0k}} \xrightarrow[n\to+\infty]{L} \chi^2_{K-1}$$

$$Z_k = \sum_{j=1}^{n} Y_j^k \text{ donc } E[Z_k] = n\, E[Y_j^k] = n\, p_{0k}$$

$$\mathrm{cov}(Z_i, Z_k) = E[Z_i Z_k] - E[Z_i]E[Z_k]$$

$$E[Z_i Z_k] = \sum_{j_1}\sum_{j_2} \underbrace{E[Y_{j_1}^k Y_{j_2}^i]}_{0 \text{ pour } j_1 \neq j_2} = \sum_j \underbrace{E[Y_j^k Y_j^i]}_{=0 \text{ pour } k\neq i}$$

donc $\boxed{\mathrm{cov}(Z_i, Z_k) = -n^2 p_{0i} p_{0k}}$

$$\mathrm{Var}\, Z_i : \mathrm{Var}\left(\sum_j Y_j^k\right) = \sum_j \mathrm{var}(Y_j^k) = n\left[\underbrace{1\, p_{0k} + 0\,(1-p_{0k})}_{E[(Y_j^k)^2]} \underbrace{- p_{0k}^2}_{-(E[Y_j^k])^2}\right]$$

$\boxed{E[Z_k] = np_{0k}}$ et $\boxed{\mathrm{Var}\, Z_k = n\, p_{0k}(1-p_{0k})}$

les variables aléatoires Z_k sont liées car $\sum_{k=1}^{K} Z_k = n$

nbre d'observations $\in$ à la classe C_k

Donc, on ne considère que

$$Z = \left(\frac{Z_1 - np_{01}}{\sqrt{n}}, \frac{Z_2 - np_{02}}{\sqrt{n}}, \dots, \frac{Z_{K-1} - np_{0K-1}}{\sqrt{n}}\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & & & \\ -np_1p_2 & p_2(1-p_2) & & \\ & & \ddots & \\ -np_1p_{K-1} & & & p_{K-1}(1-p_{K-1}) \end{pmatrix}$$

<u>1er théorème sur les formes quadratiques</u>

$$X \sim N_p(\mu, \Sigma) \Rightarrow D^2 = (X-\mu)^T \Sigma^{-1}(X-\mu) \sim \chi^2_p$$

on applique ce théorème à Z et ça marche ...

Rq 1 : $A = \sum_{k=1}^{k} \frac{n}{P_{0k}} \left(\frac{Z_k}{n} - P_{0k} \right)^2$

$\frac{Z_k}{n}$: estimation de P_{0k}

On comprend que si H_0 est vraie $\frac{Z_k}{n} \neq P_{0k}$ donc A est "petit"

On aura alors $A < S_\alpha$ et on acceptera H_0

Rq 2 : la loi de A est valable en théorie pour $n = +\infty$.

En pratique n est fini. On montre que la loi reste valable pour n fini si $n P_{0k} > 4$ ou 5 $\forall k$

On a intérêt à choisir des classes de façon à avoir P_{0k} "grand" $\forall k$

$\Rightarrow$ classes équiprobables

Rq 3 : Si L_0 est imparfaitement connue (N_p paramètres inconnus

ex : $N(m, \sigma^2)$, $N_p = 2$)

On estime m et σ^2 et $A \sim \chi^2_{k-1-N_p}$

VI Test de kolmogorov : (va continues)

Test d'ajustement : H_0 $L = L_0$

H_1 $L \neq L_0$

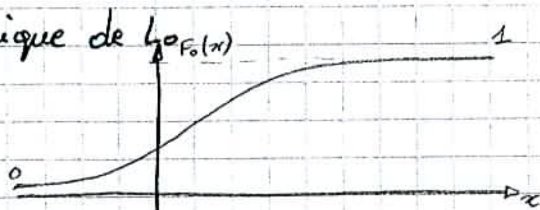
On teste si la loi des VA X_i est L_0 ou non à partir de l'observation $x_1, \dots, x_n$.

le test de kolmogorov est défini par :

$$\boxed{\text{Rejet de } H_0 \text{ si } \sup_{x \in \mathbb{R}} \left| \hat{F}(x) - F_0(x) \right| > S_\alpha}$$

où $F_0(x)$ est la f$^\circ$ de répartition théorique de L_0

$F_0(x) = \int_{-\infty}^{x} P_0(u) \, du$ (P_0 : densité de proba associée à L_0)



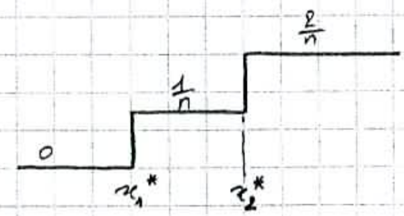
et $\hat{F}(x)$ est la f$^\circ$ de répartition empirique

$$D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}(x) - F_0(x) \right| = \sup_{i \in \{1,\dots,n\}} \left[\max \left(E_i^+, E_i^- \right) \right]$$

$$E_i^+ = \left| \hat{F}(x_i^{*+}) - F_0(x_i^*) \right|$$

$$E_i^- = \left| \hat{F}(x_i^{*-}) - F_0(x_i^*) \right|$$

<u>Si tous les x_i sont différents</u>



$$\hat{F}(x_i^{*+}) = \frac{i}{n} \quad \text{et} \quad \hat{F}(x_i^{*-}) = \frac{i-1}{n}$$

$$\boxed{D_n = \sup_i \left[\max \left| \frac{i}{n} - F_0(x_i^*) \right|, \left| \frac{i-1}{n} - F_0(x_i^*) \right| \right]}$$

<u>Rq 3:</u> • Le test de Kolmogorov est plus puissant que le test du X^2 dans la plupart des applications (puissance $= 1 - \beta$)

• Attention, son utilisation est limitée aux lois <u>continues</u>

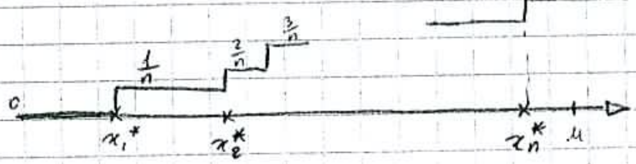
<u>Rq 4:</u> $\alpha = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$

et on rejette H_0 si $D_n > S_\alpha$, donc plus $\alpha \nearrow$

plus $S_\alpha \searrow$

et donc plus il est difficile d'accepter H_0

Si on accepte H_0, le résultat est d'autant plus fiable que α est grand



$x_1^* \leq x_2^* \leq \dots \leq x_n^*$

(On dit que $(x_1^*, \dots, x_n^*)$ est la statistique d'ordre de $(x_1, \dots, x_n)$)

<u>Rq</u>: Si tous les x_i^* sont différents deux à deux, les sauts de la fonction en escalier sont de $\frac{1}{n}$

On montre que $D_n = \sup_{x \in \mathbb{R}} |\hat{F}(x) - F_0(x)|$ possède la même loi asymptotique ($n \to \infty$) sous l'hypothèse H_0 quelle que soit la loi à tester L_0

$$P[\sqrt{n}\, D_n < y] \xrightarrow[n \to +\infty]{} \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2 y) = k(y)$$

<u>Détermination de S_α:</u>

$\alpha = P[\text{Rejeter } H_0 / H_0 \text{ vraie}]$

$\alpha = P[D_n > S_\alpha / H_0 \text{ vraie}]$

$\alpha = 1 - P[D_n \leq S_\alpha]$

$\alpha = 1 - P[\sqrt{n}\, D_n \leq \sqrt{n}\, S_\alpha]$

$\boxed{\alpha \underset{n\,gd}{\#} 1 - k(\sqrt{n}\, S_\alpha)}$

On a alors construit des tables qui donnent S_α en fonction de α

<u>Rq 1</u>: pour $n > 80$ $\quad \alpha = 0{,}05 \quad S_\alpha \# \frac{1{,}3581}{\sqrt{n}}$

$\alpha = 0{,}01 \quad S_\alpha \# \frac{1{,}6276}{\sqrt{n}}$.

<u>Rq 2</u>: <u>Calcul de D_n</u>