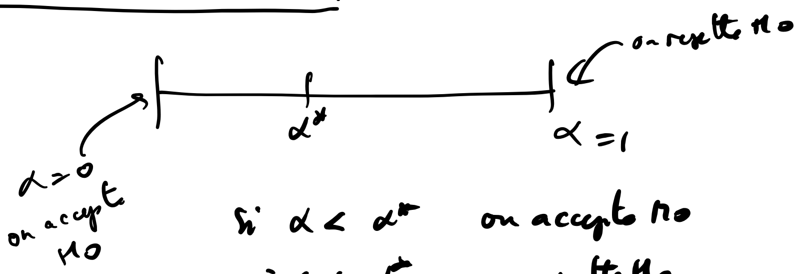
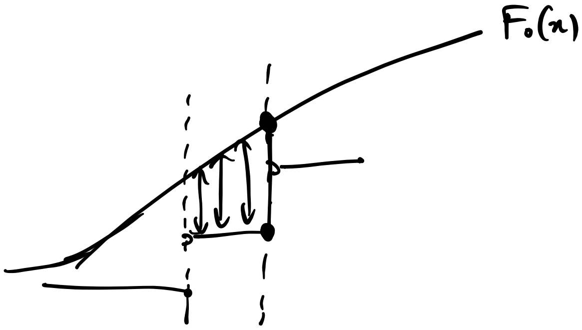
Cours du 12/01/2023

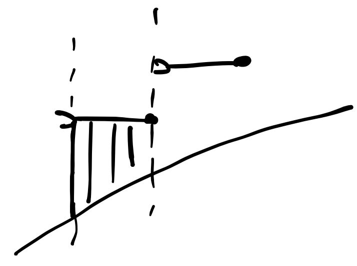


si $\alpha < \alpha^*$ on accepte H_0

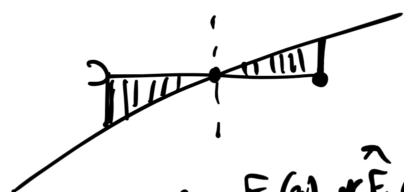
si $\alpha > \alpha^*$ on rejette H_0



<u>Cas 1</u> $F_0(x) > \hat{F}_n(x) \ \forall x$



<u>Cas 2</u> $F_0(x) < \hat{F}_n(x) \ \forall x$



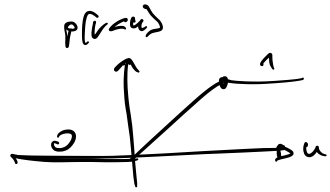
<u>Cas 3</u> : $F_0(x)$ et $\hat{F}_n(x)$ ont un point d'intersection

> Soit X une va de proba-L_0 de fonction de répartition $F_0(x) = P[X < x]$
> Quelle est la loi de $U = F_0(X)$?

U est à valeurs dans $[0,1]$

$$P[U < u] = P[F_0(X) < u] = P[X < F_0^{-1}(u)]$$

$$= F_0(F_0^{-1}(u)) = u$$



Détermination du seuil pour le test de Kolmogorov

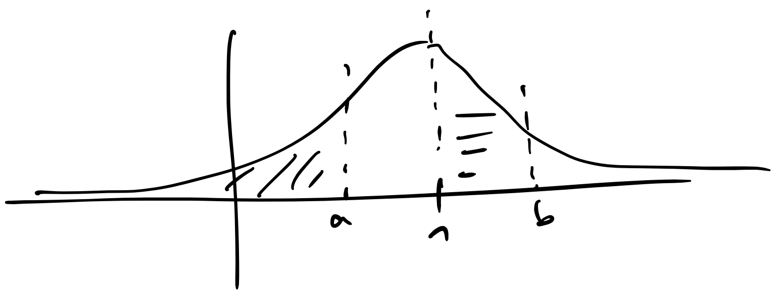
$$\alpha = P[\text{Rejeter } H_0 \mid H_0 \text{ vraie})$$

$$= P(D_n > S_\alpha \mid D_n \sim \text{Kolmogorov})$$

$$= P[\sqrt{n}\, D_n > \sqrt{n}\, S_\alpha \mid D_n \sim \text{Kolmogorov}]$$

$$= 1 - K(\sqrt{n}\, S_\alpha)$$

$$\Rightarrow \sqrt{n}\, S_\alpha = K^{-1}(1-\alpha) \Rightarrow \boxed{S_\alpha = \frac{1}{\sqrt{n}} K^{-1}(1-\alpha)}$$



$K = 4$

par défaut $K = \sqrt{n}$

a!

$$\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right] dx = 0.25$$

$m = 1$
$\sigma^2 = 4$

$u = \frac{x-m}{\sigma}$

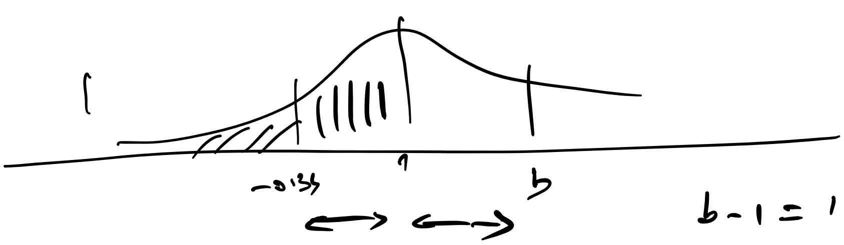
$$\int_{-\infty}^{\frac{a-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = 0.25$$

$$\frac{a-m}{\sigma} = F^{-1}(0.25)$$

$$F_{N(0,1)}\left(\frac{a-m}{\sigma}\right) = 0.25$$

$$\Rightarrow \boxed{a = m + \sigma \underbrace{F^{-1}_{N(0,1)}(0.25)}_{= -0.34}}$$

$m = 1$
$\sigma = 2$



$b - 1 = 1 - (-0.34)$

$\boxed{b = 2.34}$

$C_1 =]-\infty, -0.34]$ $\quad p_1 = \frac{1}{4}$ $\quad z_1 = 7$

$C_2 =]-0.34, 1]$ $\quad p_2 = \frac{1}{4}$ $\quad z_2 = 12$

$C_3 =]1, 2.34]$ $\quad p_3 = \frac{1}{4}$ $\quad z_3 = 10$

$C_4 =]2.34, +\infty[$ $\quad p_4 = \frac{1}{4}$ $\quad z_4 = 11$

$$\phi_n = \sum_{k=1}^{K} \frac{(z_k - np_k)^2}{np_k} \qquad np_k = 10$$

$$= \frac{(7-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(11-10)^2}{10}$$

$$= \frac{9+4+0+1}{10} = \frac{14}{10} = \boxed{1.4}$$

Sous H_0 $\quad \phi_n \sim \chi^2_{K-1} = \chi^2_{4-1} = \chi^2_3$

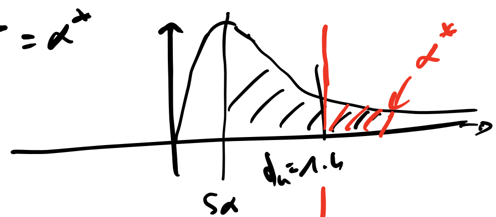
Détermination du seuil

On fixe $\alpha = 0.01 = P[\phi_n > S_\alpha \mid \phi_n \sim \chi^2_3]$

$$= 1 - F_{\chi^2_3}(S_\alpha)$$

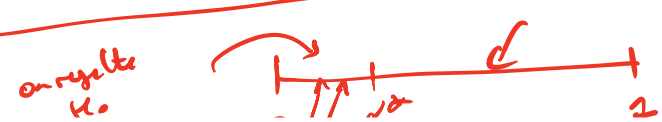
$$\Rightarrow \boxed{S_\alpha = F^{-1}_{\chi^2_3}(1-\alpha)}$$

p-valeur $= \alpha^*$



α^* = plus petite valeur de α pour laquelle on rejette H_0

$$\boxed{\alpha^* = 1 - F_{\chi^2_3}(\phi_n = 1.4)}$$



0.01 0.05

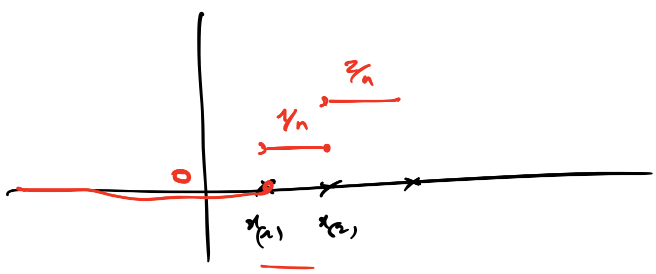
Fin cours 2 : Test χ^2 d'adéquation

TD #1 du 12 janvier

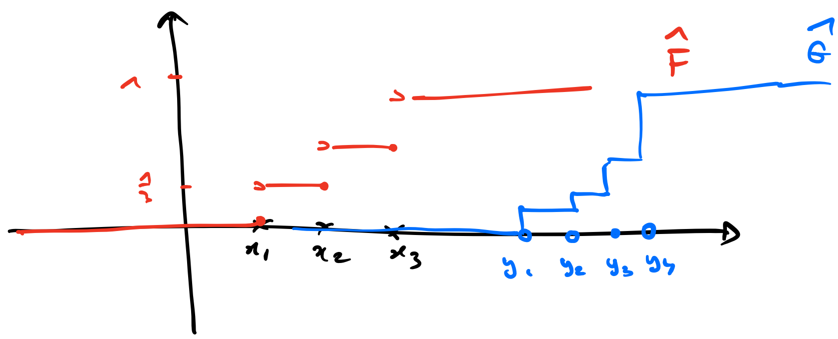
Exo 1
Exo 2 = Test de Kolmogorov Smirnov
Exo 6

Exo1

$x_{(i)}$	0.15	0.18	0.22	0.26	0.39	0.53	0.62	0.70	0.96	0.99
E_n^{i+}										
E_n^{i-}										
Max										



Cours du 13/01/2023



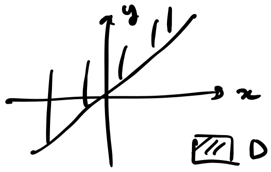
$P[Y_j > X_i] = \frac{1}{2}$!! à voir

Fin cours 3 : Intro Mann Whitney + Wilcoxon (preuve $U_Y = W_Y - \frac{m(m+1)}{2}$)
Exemple
Reste à faire : preuve des lois

$\boxed{\text{Cours 4 du 17/01/2023}}$

Rappel
$\boxed{\text{Soient } X \text{ et } Y \text{ deux variables aléatoires indépendantes et de même loi de fonction de répartition } F. \text{ Alors } P[Y > X] = \frac{1}{2}}$

$$P[Y > X] = \iint_D f(x,\cdot) f(\cdot,y)\, dx\, dy = \int_{\mathbb{R}} \left[\int_x^{+\infty} f(x,\cdot) f(\cdot,y)\, dy \right] dx$$



$$= \int_{\mathbb{R}} f(x,\cdot) \left[1 - F(x)\right] dx$$

$$= 1 - \underbrace{\int_{\mathbb{R}} F'(x) F(x)\, dx}_{\left[\frac{F^2(x)}{2}\right]_{-\infty}^{+\infty} = \frac{1}{2}} = \frac{1}{2}$$

Donc quand le nombre de couples (X_i, Y_j) (dans deux échantillons $(X_1, \dots, X_n)$ et $(Y_1, \dots, Y_m)$) qui vérifient $Y_j > X_i$ est **trop grand** ou trop petit, on a tendance à rejeter l'hypothèse H_0 : $F = G$

Rq : il faut que les 2 échantillons soient indépendants !!

Statistique de Mann-Whitney

- m, n 'grands'

$$U_Y = \sum_i \sum_j \mathbb{1}_{\{Y_j > X_i\}}$$

D'après le théorème central-limite $U_Y \sim N(E[U_Y], Var[U_Y])$ pour n et m "grands".

$$E[U_Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} \underbrace{E[\mathbb{1}_{Y_j > X_i}]}_{1 \times \underbrace{P(Y_j > X_i)}_{1/2} + 0 \times P[Y_j \leq X_i]} = \frac{nm}{2}$$

Calcul de la variance : voir slide

$$Var(U_Y) = \frac{nm(n+m+1)}{12}$$

- (m,n) "petits", on utilise la relation de récurrence

Correction de continuité

Si on fait les approximations suivantes (en invoquant central limite)

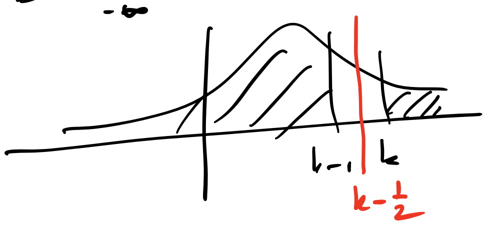
$$P[U_y \geq k] \approx \int_k^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(u-\mu)^2}{2\sigma^2}\right] du \qquad \mu = E[U_y] \quad \sigma^2 = Var[U_y]$$

et

$$P[U_y < k] = P[U_y \leq k-1] \approx \int_{-\infty}^{k-1} \text{———}\, du$$

alors $P(U_y \geq k) + P[U_y < k] = 1$

mais $$\int_k^{+\infty} + \int_{-\infty}^{k-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du \neq 1$$ car on n'a pas compté $\int_{k-1}^{k} f(u)\,du$



La correction de continuité consiste à utiliser les approximations suivantes

$$P[U_y \geq k] \approx \int_{k-\frac{1}{2}}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(u-\mu)^2}{2\sigma^2}\right] du$$

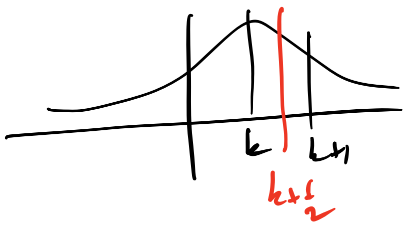
$$\approx \int_k^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(v-\frac{1}{2}-\mu)^2}{2\sigma^2}\right] du$$

$v = u + \frac{1}{2}$

Cela revient à modifier la moyenne de la loi limite $E[U_y] = \mu$ en $\boxed{E[U_y] + \frac{1}{2} = \mu + \frac{1}{2}}$

De la même façon

$$P[U_y \le k] \simeq \int_{-\infty}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left[\frac{(u-\mu)^2}{2\sigma_u^2} \right] du$$



$$\approx \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left[-\frac{1}{2\sigma_u^2} \left(v+\frac{1}{2}-\mu\right)^2 \right] du$$

$v = u - \frac{1}{2}$

Cela revient à changer $\boxed{\mu = E(U_y) \text{ en } \mu - \frac{1}{2}}$

C'est ce qu'on appelle les corrections de continuité

Droite de Henry

① Calcul de la fonction de répartition

Pour une loi normale $P[X < x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(u-\mu)^2}{2\sigma^2} \right] du$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv$$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right)$$

où $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$

donc $P[X < x_i] = \Phi\left(\frac{x_i - \mu}{\sigma}\right)$

et $P[X < x_{(i)}] = \Phi\left(\frac{x_{(i)} - \mu}{\sigma}\right)$

② on inverse la fonction de répartition et on approche $P[X < x_{(i)}]$ par $\frac{i}{n}$

$$\Phi\left(\frac{x_{(i)} - \mu}{\sigma}\right) \simeq \frac{i}{n}$$

sous l'hypothèse que les $x_{(i)}$ sont issus d'une loi normale $N(\mu, \sigma^2)$

$$\Rightarrow \frac{x_{(i)} - \mu}{\sigma} = \Phi^{-1}\left(\frac{i}{n}\right)$$

Si on pose $t_i = \phi^{-1}\left(\frac{i}{n}\right)$, alors $x_{(i)} = \mu + \sigma\, \phi^{-1}\left(\frac{i}{n}\right)$

soit $\boxed{x_{(i)} = \mu + \sigma t_i}$

la courbe $x_{(i)}$ en fonction de t_i est une droite qui s'appelle la droite de Henry

Généralisation à d'autres lois

Ex de la loi uniforme

$$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & x \in]a,b[\\ 1 & x \geq b \end{cases}$$

donc $F(x_i) = \frac{x_i - a}{b-a}$ et $F(x_{(i)}) = \frac{x_{(i)} - a}{b-a} \simeq \frac{i}{n}$

si on pose $t_i = \frac{i}{n}$, alors $\boxed{x_{(i)} \simeq (b-a)t_i + a}$ si les x_i suivent la loi uniforme. La représentation graphique des points $\left(\frac{i}{n}, x_{(i)}\right)$ doit donc être une droite si les points x_i suivent la loi uniforme