Statistiques d'ordre

$X_1, \dots, X_n$ var $\perp\!\!\!\perp \sim X$ de densité f et de fdr F.

Prop 1 : les (X_i) sont ps tous distincts.

Dem : $i \neq j$ $\mathbb{P}(X_i = X_j) = \int_{\mathbb{R}^2} \mathbb{1}_{x=y}^{(x,y)} f(x) f(y)\, dx\, dy \overset{\text{Fub}}{=} \int_{\mathbb{R}} f(x) \Big(\underbrace{\int_{\mathbb{R}} \mathbb{1}_x(y) f(y)\, dy}_{= 0 \text{ car } \{x\} \text{ est de mesure } = 0} \Big) dx$

D'où $\mathbb{P}(\exists (i \neq j), X_i = X_j) = \mathbb{P}\Big(\bigcup_{i \neq j} (X_i = X_j) \Big)$

$$\leq \sum_{i \neq j} \mathbb{P}(X_i = X_j) = 0$$

On peut donc définir pour presque tout $\omega \in \Omega$ la permutation $\sigma(\omega)$ tq
$X_{\sigma(\omega)(1)}(\omega) < \dots < X_{\sigma(\omega)(n)}(\omega)$

Prop 2 : $\sigma(\Omega) \in \mathfrak{S}_n$, $\sigma \sim \mathcal{U}(\mathfrak{S}_n)$

Dem : $\sigma^{-1}(\{\tau\})$ est mesurable pour τ permutation donc σ est une va.
Mq $\forall \tau \in \mathfrak{S}_n$ $\mathbb{P}(\sigma = \tau)$ ne dépend pas de τ pour mq $\sigma \sim \mathcal{U}(\mathfrak{S}_n)$.
$\mathbb{P}(\sigma = \tau) = \mathbb{P}(X_{\tau(1)} < \dots < X_{\tau(n)})$. Or la densité de $(X_1, \dots, X_n)$ est $f(x_1) \dots f(x_n)$
elle est invariante par permutation de coordonnées. Donc $\mathcal{L}((X_{\tau(1)}, \dots, X_{\tau(n)})) = \mathcal{L}((X_1, \dots, X_n))$
donc $\mathbb{P}(\sigma = \tau) = \mathbb{P}(X_1 < \dots < X_n)$ ne dépend pas de τ.
Donc $\sigma \sim \mathcal{U}(\mathfrak{S}_n)$.

Notation : On note $X_{(i)}(\omega) = X_{\sigma(\omega)(i)}$. En part $X_{(1)} = \min_{1 \leq i \leq n}(X_i)$
et $X_{(n)} = \max_{1 \leq i \leq n}(X_i)$
et si $n \equiv 1 [2]$ $X_{(n+1)/2}$ est la médiane des X_i.

Prop : • $(X_{(1)}, \dots, X_{(n)})$ a pour densité $n! \, \mathbb{1}_{x_1 < \dots < x_n} \prod_{i=1}^{n} f(x_i)$
• $\forall i \in [\![1, n]\!]$ $X_{(i)}$ a pour densité $i \binom{n}{i} f(x) F(x)^{i-1} (1 - F(x))^{n-i}$

Dem : • Soit $B \in \mathcal{B}(\mathbb{R}^n)$. $\mathbb{P}((X_{(1)}, \dots, X_{(n)}) \in B) = \sum_{\tau \in \mathfrak{S}_n} \mathbb{P}((X_{(1)}, \dots, X_{(n)}) \in B, \sigma = \tau)$
$= \sum_{\tau \in \mathfrak{S}_n} \mathbb{P}((X_{\tau(1)}, \dots, X_{\tau(n)}) \in B, \sigma = \tau) = n! \, \mathbb{P}((X_1, \dots, X_n) \in B \text{ et } X_1 < \dots < X_n)$
car $\mathcal{L}(X_{\tau(1)}, \dots, X_{\tau(n)}) = \mathcal{L}(X_1, \dots, X_n)$.

Et $\mathbb{P}((X_1,\dots,X_n)\in B,\ X_1<\dots<X_n) = \int_B \mathbb{1}_{x_1<\dots<x_n} f(x_1)\dots f(x_n)\,dx_1\dots dx_n$

La densité de $(X_{(1)},\dots,X_{(n)})$ est donc $n!\,\mathbb{1}_{x_1<\dots<x_n} f(x_1)\dots f(x_n)$.

• La densité de $X_{(i)}$ est donc $n!\int_{\mathbb{R}^{n-1}} \mathbb{1}_{x_1<\dots<x_n} f(x_1)\dots f(x_n)\,dx_1\dots dx_{i-1}\,dx_{i+1}\dots dx_n$

Mais on peut calculer autrement :

Soit $x\in\mathbb{R}$ $\{X_{(i)}\le x\} = \biguplus_{k\ge i}\{$"il y a k éléments du n-uplets $\le x$ et $n-k > x$"$\}$.

Soit $K\subset\{1,\dots,n\}$ tq $|K|=k$.

$\mathbb{P}(\forall j\in K\ X_j\le x \text{ et } \forall j\notin K\ X_j > x) = F(x)^k(1-F(x))^{n-k}$

Et il y a $\binom{n}{k}$ parties de cardinal k d'où

$$\mathbb{P}(X_{(i)}\le x) = \sum_{k=i}^{n} \binom{n}{k} F(x)^k(1-F(x))^{n-k}$$

Puis en dérivant : $$f_{X_{(i)}}(x) = f(x)\sum_{k=i}^{n}\binom{n}{k} kF(x)^{k-1}(1-F(x))^{n-k} - f(x)\sum_{k=i}^{n} \binom{n}{k} (n-k)F(x)^k(1-F(x))^{n-k-1}$$

$$= f(x)\sum_{k=i}^{n} \binom{n-1}{k-1} nF(x)^{k-1}(1-F(x))^{n-k} - f(x)\, n\sum_{k=i}^{n}\binom{n-1}{k}F(x)^k(1-F(x))^{n-k-1}$$

$$= nf(x)\left(\sum_{p=i-1}^{n-1}\binom{n-1}{p}F(x)^p(1-F(x))^{n-p-1} - \sum_{k=i}^{n} \binom{n-1}{k}F(x)^k(1-F(x))^{n-k-1}\right)$$

$$= n f(x)\binom{n-1}{i-1}F(x)^{i-1}(1-F(x))^{n-i} = i\binom{n}{i} f(x)F(x)^{i-1}(1-F(x))^{n-i}$$

<u>Exemples</u> : $X_i\sim\mathcal{U}([0,1])$. La densité de $(X_{(1)},\dots X_{(n)})$ est $n!\,\mathbb{1}_{x_1<\dots<x_n}$ - et celle de $X_{(i)}$ est $i\binom{n}{i} x^{i-1}(1-x)^{n-i}$ $X_{(i)}\sim\beta(i,n-i+1)$.

• $X_i\sim\mathcal{E}(1)$. La densité de $(X_{(1)},\dots,X_{(n)})$ est $n!\,\mathbb{1}_{0<x_1<\dots<x_n} e^{-\sum_{i=1}^{n} x_i}$

$Y_i = X_{(i+1)}-X_{(i)}$ a pour densité $n!\prod_{i=1}^{n} \mathbb{1}_{\mathbb{R}_+}(y_i)e^{-(n-i+1)y_i}$ (via $(x_1,\dots,x_n)\mapsto(x_1,x_2-x_1,\dots,x_n-x_{n-1})$)

et sont $\perp\!\!\!\perp$, $\sim\mathcal{E}(n-i+1)$.

$X_{(i)}$ a pour densité sur $\mathbb{R}_+$: $i\binom{n}{i} e^{-x}(1-e^{-x})^{i-1}e^{-(n-i)x} = i\binom{n}{i} e^{-(n+1-i)x}(1-e^{-x})^{i-1}$.

En part, $X_{(1)}$ a pour densité : ne^{-nx} sur $\mathbb{R}_+$ ie $X_{(1)}\sim\mathcal{E}(n)$.