Ouvrard

Statistiques d'ordre

Théorème: Soient $(X_i)_{i \leq n}$ des v.a. iid de loi p et admettant une densité f C^0_m, une fonction de répartition F. On note $r(x_1 \dots x_n) = (x_{(1)} \dots x_{(n)})$ où $x_{(1)} \leq \dots \leq x_{(n)}$ et $\{x_1 \dots x_n\} = \{x_{(1)} \dots x_{(n)}\}$ et $(X_{(1)} \dots X_{(n)}) = r(X_1 \dots X_n)$
$X_{(k)}$ est la k^e statistique d'ordre.
Alors $X_{(k)}$ admet une densité f_k.

Démonstration : ① $\{(x_1 \dots x_n) \,/\, \exists i\ x_i = x_j\}$ est une réunion finie d'hyperplans.
Cet ensemble est donc négligeable pour la mesure de Lebesgue.
Comme $(X_1 \dots X_n)$ admet une densité, $\mathbb{P}(X_i \neq X_j \text{ si } i \neq j) = 1$.

On peut donc définir la v.a. $(X_1 \dots X_n)$ sur $E = \{(x_1 \dots x_n) \in \mathbb{R}^n \,/\, \forall i \neq j,\ x_i \neq x_j\}$.
et r est continue donc mesurable sur E.
Donc $(X_{(1)} \dots X_{(n)})$ est bien une variable aléatoire

② Si $y \in \mathbb{R}$, $\mathbb{P}(X_{(k)} \leq y) = \mathbb{P}$ (k valeurs de X_i soient plus petites que y au moins)
Si $J \subset [\![1; n]\!]$, on pose $B_J = \left[\bigcap_{j \in J} (X_j \leq y)\right] \cap \left[\bigcap_{j \notin J} (X_j > y)\right]$

Donc $\mathbb{P}(X_{(k)} \leq y) = \mathbb{P}\left[\bigcup_{\substack{J \subset [\![1;n]\!] \\ |J| \geq k}} B_J\right] = \sum_{i=k}^{n} \mathbb{P}\left(\bigcup_{\substack{J \subset [\![1 \dots n]\!] \\ |J| = i}} B_J\right)$

$$\underset{\substack{\text{Si } |J| = |J'| \\ B_J \cap B_{J'} = \emptyset}}{=} \sum_{i=k}^{n} \sum_{\substack{J \subset [\![1;n]\!] \\ |J| = i}} \mathbb{P}(B_J)$$

Et si $|J| = i$, alors par indépendance des $(X_j)_{j=1}^{n}$, alors
$\mathbb{P}(B_J) = \prod_{j \in J} \mathbb{P}(X_j \leq y) \times \prod_{j \notin J} \mathbb{P}(X_j > y) = F(y)^i (1 - F(y))^{n-i}$

Donc $F_{X_{(k)}}(y) = \sum_{i=k}^{n} \binom{n}{i} F(y)^i (1 - F(y))^{n-i}$

③ Or, F est C^1_m et en chaque point de dérivabilité $F_{X_{(k)}}$ est dérivable.
La v.a. $X_{(k)}$ admet donc une densité définie presque partout et C^0 p.p.

Si F est dérivable en y,

$$f_k(y) = F'_{X_{(k)}}(y) = \sum_{i=k}^{n} \underbrace{i\binom{n}{i}}_{=\frac{n!}{(i-1)!(n-i)!} = n\binom{n-1}{i-1}} f(y)[F(y)]^{i-1}[1-F(y)]^{n-i} - \sum_{i=k}^{n-1} \underbrace{(n-i)\binom{n}{i}}_{=\frac{n!}{(n-i-1)!\,i!} = n\binom{n-1}{i}} f(y)[F(y)]^{i}[1-F(y)]^{n-i-1}$$

$$= n f(y)\left[\sum_{i=k}^{n}\binom{n-1}{i-1}F(y)^{i-1}(1-F(y))^{n-i} - \sum_{i=k}^{n-1}\binom{n-1}{i}[F(y)]^{i}[1-F(y)]^{n-i-1}\right]$$

$$f_k(y) = n f(y)\binom{n-1}{k-1}F(y)^{k-1}(1-F(y))^{n-k}$$

Théorème : Si $\mu = \mathcal{U}([0;t])$, la loi de $(X_{(1)} \cdots X_{(n)})$ est donnée par la densité
$$g(x_1 \cdots x_n) = \frac{n!}{t^n}\mathbb{1}_{(0 \leq x_1 \leq \cdots \leq x_n \leq t)}$$

Démonstration : Soit $g \in C_K^0(\mathbb{R}^n)$. $\mathbb{E}(g(X_{(1)} \cdots X_{(n)})) = \mathbb{E}(g \circ \mathcal{r}\,(X_1 \cdots X_n)) = I$

Par la formule de transfert, $I = \int_E g \circ \mathcal{r}(x_1 \cdots x_n)\, \underbrace{dP(x_1 \cdots x_n)}_{\frac{1}{t^n}\mathbb{1}_{[0;t]^n}\,dx_1 \cdots dx_n}$

Si $\sigma \in \mathfrak{S}_n$, on définit $P_\sigma : E \to E$, $(x_1 \cdots x_n) \mapsto (x_1 \cdots x_n)$ isométrie, et $E = \coprod_{\sigma \in \mathfrak{S}_n} P_\sigma^{-1}(S_n)$

où $S_n = \{x_1 \leq \cdots \leq x_n\}$.

Donc $I = \sum_{\sigma \in \mathfrak{S}_n} \int_{P_\sigma^{-1}(S_n)} g \circ \mathcal{r}(x) \times \frac{\mathbb{1}_{[0;t]^n}(x)}{t^n}\, dx.$

$$= \frac{1}{t^n}\sum_{\sigma \in \mathfrak{S}_n} \int_{\mathbb{R}^n} \mathbb{1}_{S_n}(P_\sigma(x))\, g \circ \mathcal{r}(x)\, \mathbb{1}_{[0;t]^n}(x)\, dx.$$

$y = P_\sigma(x)$

$$I = \frac{1}{t^n}\sum_{\sigma \in \mathfrak{S}_n} \int_{\mathbb{R}^n} \mathbb{1}_{S_n}(y)\, \underbrace{(g \circ \mathcal{r})(P_\sigma^{-1}(y))\, \mathbb{1}_{[0;t]^n}(P_\sigma^{-1}(y))}_{(g \circ \mathcal{r}) \cdot \mathbb{1} \text{ invariant par } P_\sigma}\, dy$$

$$I = \frac{n!}{t^n}\int_{\mathbb{R}^n} \mathbb{1}_{S_n}(y)\, g(\mathcal{r}(y))\, \mathbb{1}_{[0;t]^n}(y)\, dy = \frac{n!}{t^n}\int_{\{0 \leq x_1 \leq \cdots \leq x_n \leq t\}} g(\overbrace{\mathcal{r}(y)}^{=y})\, dy$$

Donc $I = \mathbb{E}(g(Y))$ où Y admet $\frac{n!}{t^n}\mathbb{1}_{0 \leq x_1 \leq \cdots \leq x_n \leq t}$ comme densité.

Donc $(X_{(1)} \cdots X_{(n)})$ admet $\overbrace{\frac{n!}{t^n}\mathbb{1}_{0 \leq x_1 \leq \cdots \leq x_n \leq t}}^{= f(x_1 \cdots x_n)}$ comme densité.