

Titre : Statistiques d'ordre

Recasages : 235,236,261,266

Références : Marie Cottrell, Exercices de Probabilités (p.53)

Théorème 1. Soient $X_1, \dots, X_n$ des variables aléatoires réelles, définies sur le même espace probabilisé Ω , iid de densité f et de fonction de répartition F .

Soit σ la variable aléatoire sur $\mathfrak{S}_n$ telle que, $\forall \omega \in \Omega$:

$$X_{(1)}(\omega) = X_{\sigma(\omega)(1)}(\omega) < \dots < X_{(n)}(\omega) = X_{\sigma(\omega)(n)}(\omega).$$

Alors σ est bien définie et pour tout $i \in \{1, \dots, n\}$, $X_{(i)}$ a pour densité $i \binom{n}{i} f(F)^{i-1} (1-F)^{n-i}$.

Proof. Montrons que σ est bien définie.

Si $i \neq j$, par le théorème de Fubini on a:

$$\begin{aligned} \mathbb{P}(X_i = X_j) &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x_i) f(x_j) \mathbb{1}_{\{x_i = x_j\}} dx_i dx_j \\ &= \int_{\mathbb{R}} f(x_i) \left(\int_{\{x_i\}} f(x_j) dx_j \right) dx_i = 0. \end{aligned}$$

Ainsi par sous-additivité,

$$\mathbb{P}(\exists i \neq j \mid X_i = X_j) = \mathbb{P} \left(\bigcup_{i \neq j} \{X_i = X_j\} \right) \leq \sum_{i \neq j} \mathbb{P}(X_i = X_j) = 0.$$

Les (X_i) sont p-s tous distincts et σ est (p-s) bien définie.

Montrons que σ suit une loi uniforme.

Soit $\tau \in \mathfrak{S}_n$, posons $B_\tau := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_{\tau(1)} < \dots < x_{\tau(n)}\} \in \text{Bor}(\mathbb{R}^n)$.

Comme les (X_i) sont toutes mesurables, $\sigma^{-1}(\{\tau\}) = \{\omega \in \Omega \mid (X_1(\omega), \dots, X_n(\omega)) \in B_\tau\}$ est bien un élément de la tribu de Ω , donc σ est une variable aléatoire.

Pour $\tau \in \mathfrak{S}_n$, on a :

$$\begin{aligned} \mathbb{P}(\sigma = \tau) &= \mathbb{P}(\{\omega \in \Omega \mid X_{\tau(1)}(\omega) < \dots < X_{\tau(n)}(\omega)\}) \\ &= \mathbb{P}((X_{\tau(1)}, \dots, X_{\tau(n)}) \in B_{id}) \\ &= \int_{\mathbb{R}^n} f(x_1) \dots f(x_n) \mathbb{1}_{B_{id}}(x_{\tau(1)}, \dots, x_{\tau(n)}) dx_1 \dots dx_n. \end{aligned}$$

Ainsi par le théorème de Fubini et le changement de variable $u_i = x_{\tau(i)}$ on obtient :

$$\begin{aligned} \mathbb{P}(\sigma = \tau) &= \int_{\mathbb{R}^n} f(x_{\tau(1)}) \dots f(x_{\tau(n)}) \mathbb{1}_{B_{id}}(x_{\tau(1)}, \dots, x_{\tau(n)}) dx_{\tau(1)} \dots dx_{\tau(n)} \\ &= \int_{\mathbb{R}^n} f(x_1) \dots f(x_n) \mathbb{1}_{B_{id}}(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \mathbb{P}(\sigma = id). \end{aligned}$$

Cette proba ne dépend pas de τ et alors $\sigma \sim \text{Unif}(\mathfrak{S}_n)$.

Calculons la loi de $(X_{(1)}, \dots, X_{(n)})$.

Soit $B \in \text{Bor}(\mathbb{R}^n)$, par la formule des proba totales et les calculs précédents,

$$\begin{aligned} \mathbb{P}((X_{(1)}, \dots, X_{(n)}) \in B) &= \sum_{\tau \in \mathfrak{S}_n} \mathbb{P}((X_{(1)}, \dots, X_{(n)}) \in B, \sigma = \tau) \\ &= \sum_{\tau \in \mathfrak{S}_n} \mathbb{P}((X_{\tau(1)}, \dots, X_{\tau(n)}) \in B \cap B_{id}) \\ &= \sum_{\tau \in \mathfrak{S}_n} \int_B f(x_1) \dots f(x_n) \mathbb{1}_{B_{id}}(x_{\tau(1)}, \dots, x_{\tau(n)}) dx_1 \dots dx_n \\ &= n! \int_B f(x_1) \dots f(x_n) \mathbb{1}_{B_{id}}(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

Donc $(X_{(1)}, \dots, X_{(n)})$ à pour densité

$$n! \mathbb{1}_{x_1 < \dots < x_n} \prod_{i=1}^n f(x_i).$$

Calculons la loi de $X_{(i)}$.

Pour trouver la densité de $X_{(i)}$ il suffit d'intégrer sur les autres coordonnées, on pose

$$d\hat{x}_i = dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

On obtient alors,

$$\begin{aligned} f_{X_{(i)}}(x_i) &= \int_{\mathbb{R}^{i-1} \times \mathbb{R}^{n-i}} n! f(x_1) \dots f(x_n) \mathbb{1}_{B_{id}}(x_1, \dots, x_n) d\hat{x}_i \\ &= \frac{n!}{(i-1)!(n-i)!} f(x_i) \left(\int_{]-\infty, x_i[^{i-1}} (i-1)! f(y_1) \dots f(y_{i-1}) \mathbb{1}_{y_1 < \dots < y_{i-1}} dy_1 \dots dy_{i-1} \right) \times \\ &\quad \left(\int_{]x_i, +\infty[^{n-i}} (n-i)! f(z_{i+1}) \dots f(z_n) \mathbb{1}_{z_{i+1} < \dots < z_n} dz_{i+1} \dots dz_n \right). \end{aligned}$$

Si on pose $Y_1, \dots, Y_{i-1}$ iid de densité f alors la première intégrale vaut :

$$\mathbb{P}((Y_{(1)}, \dots, Y_{(i-1)}) \in]-\infty, x_i[^{i-1}) = \mathbb{P}(Y_{(1)} < x_i, \dots, Y_{(i-1)} < x_i) = \mathbb{P}(Y_1 \leq x_i, \dots, Y_{i-1} \leq x_i).$$

Les variables étant iid cette quantité vaut $F(x_i)^{i-1}$. De la même manière on trouve que la seconde intégrale vaut $(1 - F(x_i))^{n-i}$.

Finalement

$$f_{X_{(i)}}(x_i) = i \binom{n}{i} f(x_i) F(x_i)^{i-1} (1 - F(x_i))^{n-i}.$$

□