About exchangeable random vectors

Let $(Z_1, \dots, Z_n)$ be a random vector taking its values in $\mathcal{Z}^n$.

Definition: $(Z_1, \dots, Z_n)$ is exchangeable iff, for any permutation $\sigma \in \mathfrak{S}_n$, it holds that:

$$(Z_1, \dots, Z_n) \overset{(d)}{=} (Z_{\sigma(1)}, \dots, Z_{\sigma(n)}).$$

Proposition (conditioning on the order statistics): assume that $\mathcal{Z} = \mathbb{R}$, and define the order statistics $Z_{(1)} \leq \dots \leq Z_{(n)}$ (i.e. the sorted version of the vector $(Z_1, \dots, Z_n)$).

If $(Z_1, \dots, Z_n)$ is exchangeable, it holds that:

$$(Z_1, \dots, Z_n) \mid Z_{(1)}, \dots, Z_{(n)} \sim \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \delta_{(Z_{(\sigma(1))}, \dots, Z_{(\sigma(n))})},$$

i.e. $$(Z_1, \dots, Z_n) \mid Z_{(1)} = \gamma_1, \dots, Z_{(n)} = \gamma_n \sim \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \delta_{(\gamma_{\sigma(1)}, \dots, \gamma_{\sigma(n)})}.$$

Remark: Let $(v_1, \dots, v_n) \in \mathcal{Z}^n$. When drawing $(U_1, \dots, U_n) \sim \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \delta_{(v_{\sigma(1)}, \dots, v_{\sigma(n)})}$

note that the $(v_i)_{i=1}^n$ are not necessarily unique! Thus, for any $(w_1, \dots, w_n) \in \mathcal{Z}^n$,

it leads to:

$$\mathbb{P}(U_1 = w_1, \dots, U_n = w_n) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathbb{1}\{v_{\sigma(1)} = w_1, \dots, v_{\sigma(n)} = w_n\}.$$

multiple indicator can take the value 1!

What if $\mathcal{Z} \neq \mathbb{R}$? In fact, knowing the order statistics only means knowing the values of the variables without knowing in which order they appear in the vector. This is captured in the definition of a bag / multiset.

Definition (bag / multiset): Let $(v_1, \dots, v_n) \in \mathcal{Z}^n$. The bag (or multiset) $\{\!\{ v_1, \dots, v_n \}\!\}$ contains all the elements of $(v_1, \dots, v_n)$ unordered.

It is an object of size n: it contains duplicates when there are in $(v_1, \dots, v_n)$.

For a bag $\{\!\{ v_1, \dots, v_n \}\!\}$, we define

$$\varphi_{\{\!\{ v_1, \dots, v_n \}\!\}} : \mathcal{Z}^n \longrightarrow \mathbb{N}$$
$$(w_1, \dots, w_n) \longmapsto \#\{\sigma \in \mathfrak{S}_n : v_{\sigma(1)} = w_1, \dots, v_{\sigma(n)} = w_n\}$$
$$= \sum_{\sigma \in \mathfrak{S}_n} \mathbb{1}\{v_{\sigma(1)} = w_1, \dots, v_{\sigma(n)} = w_n\}$$

Ex.: * for $(v_1, \dots, v_n) = (1, 3, 2, 3)$, we have $\{\!\{ v_1, \dots, v_n \}\!\} = \{\!\{ 1, 2, 3, 3 \}\!\}$

* $\{\!\{ 1, 2, 3, 3 \}\!\} = \{\!\{ 3, 2, 1, 3 \}\!\} \neq \{\!\{ 1, 3, 2 \}\!\}$

* $\varphi_{\{\!\{1,2,3,3\}\!\}}((1,3,2,3)) = 2 \qquad \varphi_{\{\!\{1,2,3,4\}\!\}}((4,2,1,3)) = 1 \qquad \varphi_{\{\!\{1,2\}\!\}}((3,3)) = 0$

Proposition (conditioning on the bag): If $(Z_1, \dots, Z_n)$ is exchangeable, it holds that:

$$(Z_1, \dots, Z_n) \mid \{\!\{ Z_1, \dots, Z_n \}\!\} \sim \underbrace{\frac{1}{n!} \sum_{\sigma \in S_n} \delta_{(Z_{\sigma(1)}, \dots, Z_{\sigma(n)})}}_{\mathcal{U}(\{(Z_{\sigma(1)}, \dots, Z_{\sigma(n)}), \sigma \in S_n\})} .$$

Proof: First, remark that $\operatorname{supp}((Z_1, \dots, Z_n) \mid \{\!\{ Z_1, \dots, Z_n \}\!\}) = \{(Z_{\sigma(1)}, \dots, Z_{\sigma(n)}), \sigma \in S_n\}$.

Thus, let $(\omega_1, \dots, \omega_n) \in \{(Z_{\sigma(1)}, \dots, Z_{\sigma(n)}), \sigma \in S_n\}$.

To prove the result, we are going to use Bayes theorem. $P(A \mid B) = \frac{P(A, B)}{P(B)}$

Thus, let $R \subseteq \{\{\!\{ z_1, \dots, z_n \}\!\}, (z_1, \dots, z_n) \in \mathcal{Z}^n\}$.

$$\mathbb{P}(Z_1 = \omega_1, \dots, Z_n = \omega_n \mid \{\!\{ Z_1, \dots, Z_n \}\!\} \in R)$$

$$= \frac{\mathbb{P}(Z_1 = \omega_1, \dots, Z_n = \omega_n, \{\!\{ Z_1, \dots, Z_n \}\!\} \in R)}{\mathbb{P}(\{\!\{ Z_1, \dots, Z_n \}\!\} \in R)}$$

by definition of exchangeability $\rightarrow$ $$= \frac{\mathbb{P}(Z_{\sigma(1)} = \omega_1, \dots, Z_{\sigma(n)} = \omega_n, \{\!\{ Z_{\sigma(1)}, \dots, Z_{\sigma(n)} \}\!\} \in R)}{\mathbb{P}(\{\!\{ Z_1, \dots, Z_n \}\!\} \in R)} \quad \text{for any } \sigma \in S_n$$

bags are permutation-invariant $\{\!\{ Z_1, \dots, Z_n \}\!\} = \{\!\{ Z_{\sigma(1)}, \dots, Z_{\sigma(n)} \}\!\}$ for any $\sigma \in S_n$ $\rightarrow$ $$= \frac{\mathbb{P}(Z_{\sigma(1)} = \omega_1, \dots, Z_{\sigma(n)} = \omega_n, \{\!\{ Z_1, \dots, Z_n \}\!\} \in R)}{\mathbb{P}(\{\!\{ Z_1, \dots, Z_n \}\!\} \in R)} \quad \text{for any } \sigma \in S_n$$

$$= \frac{1}{n!} \sum_{\sigma \in S_n} \mathbb{P}(Z_{\sigma(1)} = \omega_1, \dots, Z_{\sigma(n)} = \omega_n \mid \{\!\{ Z_1, \dots, Z_n \}\!\} \in R)$$

$$= \frac{1}{n!} \mathbb{E}\left[\sum_{\sigma \in S_n} \mathbb{1}\{Z_{\sigma(1)} = \omega_1, \dots, Z_{\sigma(n)} = \omega_n\} \mid \{\!\{ Z_1, \dots, Z_n \}\!\} \in R \right]$$

$$= \frac{1}{n!} \varphi_{\{\!\{ Z_1, \dots, Z_n \}\!\}}(\omega_1, \dots, \omega_n)$$

$\square$

We now understand that the only requirement for the result to hold, other than exchangeability, is to condition on a random variable that is permutation-invariant. Hence, we have the following generalized result:

Proposition (conditioning on any permutation-invariant function): Let g be a function with argument in $\mathcal{Z}^n$,

that is permutation-invariant (i.e. $g(z_1, \dots, z_n) = g(z_{\sigma(1)}, \dots, z_{\sigma(n)})$ for any $\sigma \in S_n$, $(z_1, \dots, z_n) \in \mathcal{Z}^n$).

Then, if $(Z_1, \dots, Z_n)$ is exchangeable, it holds that:

$$(Z_1, \dots, Z_n) \mid g(Z_1, \dots, Z_n) \sim \mathcal{U}\left(\{(Z_{\sigma(1)}, \dots, Z_{\sigma(n)}), \sigma \in S_n\}\right).$$

☆

Corollary (conditioning on the empirical measure): Denote $\hat{D}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{Z_i}$ the empirical distribution.

If $(Z_1, \dots, Z_n)$ is exchangeable, it holds that:

$$(Z_1, \dots, Z_n) \mid \hat{D}_n \sim \mathcal{U}\left(\{(Z_{\sigma(1)}, \dots, Z_{\sigma(n)}), \sigma \in S_n\}\right).$$

☆

Given the characterization of the conditional distribution of $(Z_1, \dots, Z_n)$ in the exchangeable setting, we can deduce the conditional distribution of each Z_i, $i \in [\![1, n]\!]$.

Proposition: Let g be a permutation-invariant function with argument in $\mathcal{Z}^n$.

e.g. order statistics, bag, empirical measure ...

If $(Z_1, \dots, Z_n)$ is exchangeable, for any $i \in [\![1, n]\!]$ we have:

$$Z_i \mid g(Z_1, \dots, Z_n) \sim \mathcal{U}\left(\{Z_1, \dots, Z_n\}\right) = \frac{1}{n}\sum_{i=1}^{n} \delta_{Z_i}.$$

☆

Corollary (quantile lemma): If $(Z_1, \dots, Z_n)$ is exchangeable, for any $i \in [\![1, n]\!]$

and $\beta \in [0, 1]$, we have:

$$\mathbb{P}\left(Z_i \leq q_\beta(Z_1, \dots, Z_n)\right) \in \left[\beta ; \beta + \tfrac{1}{n}\right[,$$

where the upper bound holds if the $(Z_i)_{i=1}^n$ are a.s. distinct.

☆

Proof: Let $\beta \in (0,1)$ and let $i \in [\![1,n]\!]$

$$\mathbb{P}(Z_i \leq q_\beta(Z_1,\dots,Z_n)) = \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\{Z_i \leq q_\beta(Z_1,\dots,Z_n)\} \mid q_\beta(Z_1,\dots,Z_n)\right]\right]$$

by the previous proposition $\Rightarrow$ $= \mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\{Z_j \leq q_\beta(Z_1,\dots,Z_n)\} \mid q_\beta(Z_1,\dots,Z_n)\right]\right]$ for any j

$$= \mathbb{P}(Z_j \leq q_\beta(Z_1,\dots,Z_n)) \quad \text{for any } j \in [\![1,n]\!]$$

$$= \frac{1}{n}\sum_{j=1}^{n} \mathbb{P}(Z_j \leq q_\beta(Z_1,\dots,Z_n))$$

$$= \frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{1}\{Z_i \leq q_\beta(Z_1,\dots,Z_n)\}\right]$$

$$\geq \frac{1}{n}\mathbb{E}\left[\lceil \beta n \rceil\right] \geq \beta$$

if the $(Z_i)_{i=1}^n$ are a.s. distinct,

then it becomes an equality : $= \frac{1}{n}\mathbb{E}\left[\lceil \beta n \rceil\right] < \beta + 1/n$

$\square$