ROTATION SAMPLING FOR FUNCTIONAL DATA

David Degras

DePaul University

Supplementary Material

S1 Proof of Proposition 1

It suffices to prove the proposition at each replacement time τ_r, $r = 0, \dots, m$. To do so, we proceed by induction on r. For $r = 0$, the property is true by definition of SRSWOR. Assume that the property holds at rank $(r-1)$ for some $0 < r < m$. Fix the stratum U_h and consider a subset $D \subset U_h$ of size $n_h(\tau_r)$. In order to establish the property at rank r, we must show that

$$P\left(s_h(\tau_r) = D\right) = \binom{N_h}{n_h(\tau_r)}^{-1}. \tag{S1.1}$$

By the total probability formula and the induction assumption,

$$\begin{aligned} P\left(s_h(\tau_r) = D\right) &= \sum_{\substack{D' \subset U_h \\ \#(D') = n_h(\tau_{r-1})}} P\left(s_h(\tau_r) = D \middle| s_h(\tau_{r-1}) = D'\right) P\left(s_h(\tau_{r-1}) = D'\right) \\ &= \binom{N_h}{n_h(\tau_{r-1})}^{-1} \sum_{\substack{D' \subset U_h \\ \#(D') = n_h(\tau_{r-1})}} P\left(s_h(\tau_r) = D \middle| s_h(\tau_{r-1}) = D'\right). \end{aligned} \tag{S1.2}$$

We seek the subsets $D' \subset U_h$ of size $n_h(\tau_{r-1})$ such that $P\left(s_h(\tau_r) = D \middle| s_h(\tau_{r-1}) = D'\right) > 0$. Let D' be such a subset and let $k = \#(D \cap D')$. For the sample $s_h(\tau_{r-1}) = D'$ to transform into $s_h(\tau_r) = D$, the $(n_h(\tau_{r-1}) - k)$ units in $D' \setminus D$ must be removed from $s_h(\tau_{r-1})$ and the $(n_h(\tau_r) - k)$ units in $D \setminus D'$ must be added to $s_h(\tau_{r-1})$. This entails that $k = (1-\alpha_h)n_h(\tau_{r-1})$. Reciprocally, any subset $D' \subset U_h$ of size $n_h(\tau_{r-1})$ verifying the condition $\#(D \cap D') = (1-\alpha_h)n_h(\tau_{r-1})$ can be transformed in D with the above operations. This condition is thus necessary and sufficient and the number $d_h(\tau_r)$ of subsets $D' \subset U_h$ of size $n_h(\tau_{r-1})$ satisfying it is

$$d_h(\tau_r) = \binom{n_h(\tau_r)}{(1-\alpha_h)n_h(\tau_{r-1})} \binom{N_h - n_h(\tau_r)}{n_h(\tau_{r-1}) - (1-\alpha_h)n_h(\tau_{r-1})}, \tag{S1.3}$$

where the first factor accounts for the possible choices of the $(1-\alpha_h)n_h(\tau_{r-1})$ common elements between D and D' and the second factor accounts for the possible choices of the $(n_h(\tau_r)-(1-\alpha_h)n_h(\tau_{r-1}))$ remaining elements of D' in $U_h \setminus D$).

For each of the previous subsets, the properties of SRSWOR imply that

$$P\left(s_h(\tau_r)=D \middle| s_h(\tau_{r-1})=D'\right) = \left[\binom{N_h - n_h(\tau_{r-1})}{n_h(\tau_r)-(1-\alpha_h)n_h(\tau_{r-1})}\binom{n_h(\tau_{r-1})}{\alpha_h n_h(\tau_{r-1})}\right]^{-1}. \quad \text{(S1.4)}$$

Plugging (S1.3)-(S1.4) in (S1.2), one deduces (S1.1), which completes the induction. □

S2 Proof of Lemma 1

Let $D_0,\dots,D_m \subset D$. Fix $r\in\{1,\dots,m\}$. We will establish that

$$\begin{aligned}&P\left(s_h(\tau_r)\cap D = D_r \middle| s_h(\tau_{r-1})\cap D = D_{r-1},\dots,s_h(\tau_0)\cap D = D_0\right)\\ &\qquad = P\left(s_h(\tau_r)\cap D = D_r \middle| s_h(\tau_{r-1})\cap D = D_{r-1}\right).\end{aligned} \quad \text{(S2.1)}$$

For $1\le i\le r$, we express $\{s_h(\tau_i)\cap D = D_i\}$ more conveniently as $\{s_h(\tau_i)\in A_i\}$, where $A_i = \{a_i \in \mathcal{P}(U_h) : \#a_i = n_h(\tau_i), a_i \supset D_i\}$. ($\mathcal{P}(U_h)$ is the set of all subsets of U_h). We denote by $\mathbf{a}$ the generic element $(a_{r-1},\dots,a_0)\in A_{r-1}\times\cdots\times A_0$. Let $A=\{\mathbf{a}\in A_{r-1}\times\cdots\times A_0 : P(\mathbf{a})>0\}$. The Markov property of $\{s_h(\tau_0),\dots,s_h(\tau_m)\}$ yields

$$\begin{aligned}&P\left(s_h(\tau_r)\in A_r \middle| s_h(\tau_{r-1})\in A_{r-1},\dots,s_h(\tau_0)\in A_0\right)\\ &= \frac{\sum_{\mathbf{a}\in A} P\left(s_h(\tau_r)\in A_r \middle| s_h(\tau_{r-1})=a_{r-1}\right) P\left(s_h(\tau_{r-1})=a_{r-1},\dots,s_h(\tau_0)=a_0\right)}{\sum_{\mathbf{a}\in A} P\left(s_h(\tau_{r-1})=a_{r-1},\dots,s_h(\tau_0)=a_0\right)}.\end{aligned} \quad \text{(S2.2)}$$

Invoking again the Markov property, we obtain that

$$\begin{aligned}&P\left(s_h(\tau_{r-1})=a_{r-1},\dots,s_h(\tau_0)=a_0\right)\\ &\quad = P\left(s_h(\tau_{r-1})=a_{r-1} \middle| s_h(\tau_{r-2})=a_{r-2}\right)\times\cdots\\ &\qquad \times P\left(s_h(\tau_1)=a_1 \middle| s_h(\tau_0)=a_0\right)\times P\left(s_h(\tau_0)=a_0\right)\end{aligned} \quad \text{(S2.3)}$$

for all $\mathbf{a}\in A$.

Equation (S1.4) and the properties of SRSWOR show that (S2.3) only depends on α_h, $n_h(\tau_0),\dots,n_h(\tau_r)$ and N_h. As a consequence, (S2.2) rewrites as

$$\begin{aligned}&P\left(s_h(\tau_r)\in A_r \middle| s_h(\tau_{r-1})\in A_{r-1},\dots,s_h(\tau_0)\in A_0\right)\\ &\qquad = \frac{1}{\#A}\sum_{\mathbf{a}\in A} P\left(s_h(\tau_r)\in A_r \middle| s_h(\tau_{r-1})=a_{r-1}\right).\end{aligned} \quad \text{(S2.4)}$$

Each collection A_i describes the value of the sample $s_h(\tau_i)$ in D, namely D_i, and the unspecified possible values of $s_h(\tau_i)$ outside of D. In view of Proposition 1, it is easily seen that the conditional probability distribution of $s_h(\tau_{r-1}) \setminus D$ given that $s_h(\tau_{r-1}) \cap D = D_{r-1}$ coincides with the SRSWOR of $n_h(\tau_{r-1}) - d_{r-1}$ units in $U_h \setminus D$. A combinatorial consequence is that

$$\#\big\{(a_{r-2}, \dots, a_0) \in A_{r-2} \times \dots \times A_0 : (a_{r-1}, \dots, a_0) \in A\big\} = \frac{\#A}{\#A_{r-1}} \tag{S2.5}$$

for all $a_{r-1} \in A_{r-1}$. By following the proof of Proposition 1, one could explicitly find $(\#A/\#A_{r-1})$ in terms of α_h, N_h and the $n_h(\tau_i), d_i, \#(D_i \cap D_{i+1})$, $0 \le i \le r-1$.

Combining (S2.4) and (S2.5), we finally obtain

$$\begin{aligned} &P\left(s_h(\tau_r) \in A_r \middle| s_h(\tau_{r-1}) \in A_{r-1}, \dots, s_h(\tau_0) \in A_0\right) \\ &\qquad = \frac{1}{\#A_{r-1}} \sum_{a_{r-1} \in A_{r-1}} P\left(s_h(\tau_r) \in A_r \middle| s_h(\tau_{r-1}) = a_{r-1}\right) \\ &\qquad = P\left(s_h(\tau_r) \in A_r \middle| s_h(\tau_{r-1}) \in A_{r-1}\right). \quad \square \end{aligned}$$

S3 Proof of Proposition 2

We decompose the studied sum as $\sum_{\ell=1}^{4} A_\ell(t,t')$, where

$$A_\ell(t,t') = \sum_{\substack{i,j,k,l \in U_h \\ \mathcal{C}_{ijkl}=\ell}} \mathbb{E}\left(I_i(t)I_j(t)I_k(t')I_l(t')\right) \tilde{X}_i(t)\tilde{X}_j(t)\tilde{X}_k(t')\tilde{X}_l(t')$$

and $\mathcal{C}_{ijkl} = \#\{i,j,k,l\}$. Hereafter, we compute $\mathbb{E}\left(I_i(t)I_j(t)I_k(t')I_l(t')\right)$ using the properties of SRSWOR and we compute sums involving $\tilde{X}_i(t)\tilde{X}_j(t)\tilde{X}_k(t')\tilde{X}_l(t')$ using the identity $\sum_{k \in U_h} \tilde{X}_k(t) = 0$. Let i^*, j^*, j^*, l^* be four distinct units in U_h.

We begin with the straightforward calculation of $A_1(t,t')$:

$$A_1(t,t') = \mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')\right) \sum_k \tilde{X}_k^2(t)\tilde{X}_k^2(t'). \tag{S3.1}$$

The term $A_2(t,t')$ can be expressed as

$$\begin{aligned} A_2(t,t') = {} & \mathbb{E}\left(I_{i^*}(t)I_{k^*}(t')\right) \sum_{i \ne k} \tilde{X}_i^2(t)\tilde{X}_k^2(t') \\ & + 2\,\mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')I_{k^*}(t)I_{k^*}(t')\right) \sum_{i \ne l} \tilde{X}_i(t)\tilde{X}_i(t')\tilde{X}_l(t)\tilde{X}_l(t') \\ & + 2\,\mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')I_{k^*}(t')\right) \sum_{i \ne k} \tilde{X}_i^2(t)\tilde{X}_i(t')\tilde{X}_k(t') \\ & + 2\,\mathbb{E}\left(I_{i^*}(t)I_{k^*}(t)I_{k^*}(t')\right) \sum_{i \ne k} \tilde{X}_i(t)\tilde{X}_k(t)\tilde{X}_k^2(t'), \end{aligned}$$

that is,

$$\begin{aligned} A_2(t,t') &= \mathbb{E}\left(I_{i^*}(t)I_{k^*}(t')\right)\left[(N_h-1)^2\gamma_h(t,t)\gamma_h(t',t') - \sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t')\right] \\ &\quad + 2\,\mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')I_{k^*}(t)I_{k^*}(t')\right)\left[(N_h-1)^2\gamma_h^2(t,t') - \sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t')\right] \\ &\quad - 2\left[\mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')I_{k^*}(t')\right) + \mathbb{E}\left(I_{i^*}(t)I_{k^*}(t)I_{k^*}(t')\right)\right]\sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t'). \end{aligned} \tag{S3.2}$$

Next, we have

$$\begin{aligned} A_3(t,t') &= \mathbb{E}\left(I_{i^*}(t)I_{j^*}(t)I_{k^*}(t')\right)\sum_{i\neq j\neq k}\tilde{X}_i(t)\tilde{X}_j(t)\tilde{X}_k^2(t') \\ &\quad + \mathbb{E}\left(I_{i^*}(t)I_{k^*}(t')I_{l^*}(t')\right)\sum_{i\neq k\neq l}\tilde{X}_i^2(t)\tilde{X}_k(t')\tilde{X}_l(t') \\ &\quad + 4\,\mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')I_{j^*}(t)I_{k^*}(t')\right)\sum_{i\neq j\neq k}\tilde{X}_{i^*}(t)\tilde{X}_{i^*}(t)\tilde{X}_j(t)\tilde{X}_k(t') \end{aligned}$$

and a further expansion yields

$$\begin{aligned} A_3(t,t') &= \left[\mathbb{E}\left(I_{i^*}(t)I_{j^*}(t)I_{k^*}(t')\right) + \mathbb{E}\left(I_{i^*}(t)I_{k^*}(t')I_{l^*}(t')\right)\right] \\ &\qquad \times\left[-\left(N_h-1\right)^2\gamma_h(t,t)\gamma_h(t',t') + 2\sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t')\right] \\ &\quad + 4\,\mathbb{E}\left(I_{i^*}(t)I_{i^*}(t')I_{j^*}(t)I_{k^*}(t')\right) \\ &\qquad \times\left[-\left(N_h-1\right)^2\gamma_h^2(t,t') + 2\sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t')\right]. \end{aligned} \tag{S3.3}$$

To compute $A_4(t,t')$, recall that $\sum_{i,j,k,l\in U_h}\tilde{X}_i(t)\tilde{X}_j(t)\tilde{X}_k(t')\tilde{X}_l(t') = 0$ and use the decomposition

$$\sum_{i,j,k,l\in U_h}\tilde{X}_i(t)\tilde{X}_j(t)\tilde{X}_k(t')\tilde{X}_l(t') = \sum_{\ell=1}^{4}\sum_{\substack{i,j,k,l\in U_h\\ \mathcal{C}_{ijkl}=\ell}}\tilde{X}_i(t)\tilde{X}_j(t)\tilde{X}_k(t')\tilde{X}_l(t')$$

together with the expressions of $A_1(t,t'), A_2(t,t'), A_3(t,t')$ to obtain

$$\begin{aligned} &A_4(t,t') = \mathbb{E}\left(I_{i^*}(t)I_{j^*}(t)I_{k^*}(t')I_{l^*}(t')\right)\times \\ &\left[\left(N_h-1\right)^2\gamma_h(t,t)\gamma_h(t',t') + 2\left(N_h-1\right)^2\gamma_h^2(t,t') - 6\sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t')\right]. \end{aligned} \tag{S3.4}$$

The proof is completed by gathering (S3.1)–(S3.4) and observing that all terms involving $\sum_{k\in U_h}\tilde{X}_k^2(t)\tilde{X}_k^2(t')$ are of lower order N_h thanks to (A1). $\square$