Noe's Recursion

Marc Noé (1972). The calculation of distributions of two-sided Kolmogorov–Smirnov type statistics. *Ann. Math. Statist.* ***43****(1)*, 58–64.

Let $\xi_1,\ \xi_2,\ \xi_3,\ \ldots$ be stochastically independent random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with uniform distribution on $[0, 1]$. For an integer $n \geq 1$ let $\xi_{n:1} < \xi_{n:2} < \cdots < \xi_{n:n}$ be the order statistics of $\xi_1, \xi_2, \ldots, \xi_n$. The task is to compute

$$Q_n \ := \ \mathbb{P}\big(\xi_{n:i} \in (\alpha_i, \beta_i] \text{ for } 1 \leq i \leq n\big)$$

for given numbers $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n$ and $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$ in $[0, 1]$, where $\alpha_i < \beta_i$ for $1 \leq i \leq n$.

For integers $0 \leq k \leq \ell \leq n$ and real numbers $0 \leq r \leq t \leq 1$ let

$$P_{k,\ell}(r, t) \ := \ \mathbb{P}(B_{k,\ell}(r, t)),$$

where

$$B_{k,\ell} \ := \ \begin{cases} \Omega & \text{if } k = \ell, \\ \big[\xi_{\ell-k:i} \in (\alpha_{k+i}, \beta_{k+i}] \cap (r, t] \text{ for } 1 \leq i \leq \ell - k\big] & \text{if } k < \ell. \end{cases}$$

That is, $Q_n = P_{0,n}(0, 1)$. Now,

$$P_{0,\ell}(0, t) \ = \ \sum_{k=0}^{\ell} \binom{\ell}{k} P_{0,k}(0, r) P_{k,\ell}(r, t). \tag{1}$$

Indeed, the event $B_{0,\ell}(0, t)$ is the union of the pairwise disjoint events

$$B_{0,\ell}(0, t) \cap [\xi_{\ell:k} \leq r < \xi_{\ell:k+1}], \quad 0 \leq k \leq \ell,$$

where $\xi_{\ell:0} := 0$ and $\xi_{\ell:\ell+1} := 1$. For fixed integers $0 \leq k \leq \ell$, the event $[\xi_{\ell:k} \leq r < \xi_{\ell:k+1}]$ is the union of the $\binom{\ell}{k}$ pairwise disjoint events

$$A_I \ := \ \big[\xi_i \leq r \text{ for } i \in I,\ \xi_i > r \text{ for } i \in \{1, \ldots, \ell\} \setminus I\big],$$

where $I \subset \{1, \ldots, \ell\}$ with $\#I = k$. Since $(\xi_i)_{i \in I}$ and $(\xi_i)_{i \in \{1,\ldots,\ell\} \setminus I}$ are stochastically independent and have the same distribution as $(\xi_i)_{i \leq k}$ and $(\xi_i)_{i \leq \ell - k}$, respectively,

$$\begin{aligned} \mathbb{P}(B_{0,\ell}(0, t) \cap A_I) \ &= \mathbb{P}\big(\xi_{k:i} \in (\alpha_i, \beta_i] \cap (0, r] \text{ for } 1 \leq i \leq k\big) \\ &\quad \cdot \mathbb{P}\big(\xi_{\ell-k:i} \in (\alpha_{k+i}, \beta_{k+i}] \cap (r, t] \text{ for } 1 \leq i \leq \ell - k\big) \\ &= P_{0,k}(0, r) P_{k,\ell}(r, t). \end{aligned}$$

Now let $\gamma_0 < \gamma_1 < \cdots < \gamma_N$ be the different elements of $\{\alpha_1, \ldots, \alpha_n\} \cup \{\beta_1, \ldots, \beta_N\}$, that is, $N < 2n$, $\gamma_0 = \alpha_1$ and $\gamma_N = \beta_n$. For $\ell \in \{1, \ldots, m\}$ and $m \in \{1, \ldots, N\}$ define

$$Q_{\ell,m} \ := \ P_{0,\ell}(\gamma_0, \gamma_m) \ = \ \mathbb{P}\big(\xi_{\ell:i} \in (\alpha_i, \beta_i] \cap (0, \gamma_m] \text{ for } 1 \leq i \leq \ell\big),$$

so

$$Q_n \;=\; Q_{n,N}.$$

For $m \in \{1,\ldots,N\}$ let

$$\begin{aligned} k(m) &:= \min\bigl\{k \in \{0,\ldots,n-1\}\colon \beta_{k+1} \geq \gamma_m\bigr\}, \\ \ell(m) &:= \max\bigl\{\ell \in \{1,\ldots,n\}\colon \alpha_\ell < \gamma_m\bigr\}. \end{aligned}$$

These numbers are well-defined, because $\beta_n = \gamma_N \geq \gamma_m > \gamma_0 = \alpha_1$. Then, for $1 \leq \ell \leq n$ and $2 \leq m \leq N$,

$$\begin{aligned} Q_{\ell,1} &= \begin{cases} (\gamma_1 - \gamma_0)^\ell & \text{if } \ell \leq \ell(1), \\ 0 & \text{if } \ell > \ell(1), \end{cases} \\ Q_{\ell,m} &= \begin{cases} \sum_{k=k(m)}^{\ell} \binom{\ell}{k} Q_{k,m-1} (\gamma_m - \gamma_{m-1})^{\ell-k} & \text{if } k(m) < \ell \leq \ell(m), \\ Q_{\ell,m-1} & \text{if } \ell \leq \min\{\ell(m), k(m)\}, \\ 0 & \text{if } \ell > \ell(m), \end{cases} \end{aligned}$$

where $Q_{0,m} := 1$. The formula for $Q_{\ell,1}$ follows from the fact that $\beta_i \geq \beta_1 \geq \gamma_1$ for all i, whence

$$Q_{\ell,1} \;=\; 1_{\alpha_\ell = \gamma_0]}\, \mathbb{P}(\xi \in (\gamma_0, \gamma_1] \text{ for } 1 \leq i \leq \ell) =\; 1_{[\ell \leq \ell(1)]} (\gamma_1 - \gamma_0)^\ell.$$

As to the formula for $Q_{\ell,m}$, $m \geq 2$, it follows from recursion (**??**) that

$$Q_{\ell,m} \;=\; \sum_{k=0}^{\ell} \binom{\ell}{k} Q_{k,m-1} P_{k,\ell}(\gamma_{m-1}, \gamma_m),$$

and $P_{\ell,\ell}(\gamma_{m-1}, \gamma_m) = 1 = (\gamma_m - \gamma_{m-1})^0$, while for $0 \leq k < \ell$,

$$\begin{aligned} P_{k,\ell}(\gamma_{m-1}, \gamma_m) &= \mathbb{P}\bigl(\xi_{\ell-k:i} \in (\alpha_{k+i}, \beta_{k+i}] \cap (\gamma_{m-1}, \gamma_m] \text{ for } 1 \leq i \leq \ell - k\bigr) \\ &= \begin{cases} \mathbb{P}(\xi_{\ell-k:i} \in (\gamma_{m-1}, \gamma_m] \text{ for } 1 \leq i \leq \ell - k) & \text{if } \beta_{k+1} \geq \gamma_m > \alpha_\ell, \\ 0 & \text{else}, \end{cases} \\ &= \begin{cases} (\gamma_m - \gamma_{m-1})^{\ell-k} & \text{if } k \geq k(m) \text{ and } \ell \leq \ell(m), \\ 0 & \text{else}. \end{cases} \end{aligned}$$

Remark 1. If computation of $Q_n = Q_{n,N}$ is the only goal, it suffices to compute $Q_{\ell,m}$ recursively for $m = 1, 2, \ldots, N$ and $k(m) \leq \ell \leq \ell(m)$, because $Q_{n,N}$ does not require any value $Q_{\ell,m}$ with $\ell \leq k(m)$.

Remark 2. Numerical experiments show that for large values n, say, $n \geq 1000$, the computed values of Q_n become affected by rounding errors. The main problem seem to be very small factors $\gamma_m - \gamma_{m-1}$ or very small probabilities $Q_{\ell,m}$. Apparently, these problems can be alleviated by working with $\log Q_{\ell,m} \in [-\infty, 0]$ and using the formula

$$\log(a + b) \;=\; \max\{\log a, \log b\} + \log\bigl(1 + \exp(-|\log a - \log b|)\bigr)$$

for $a, b \geq 0$ with $a + b > 0$.