Survey on approximation methods for discrete scan statistics: a software illustration

Alexandru Amărioarei, Cristian Preda

Laboratoire de Mathématiques Paul Painlevé
Département de Probabilités et Statistique
Université de Lille 1, INRIA/Modal Team, France

International Workshop on Applied Probability
16-19 June, 2014, Antalya, Turkey

Outline

Outline

The d-dimensional discrete scan statistics

Let $T_1, T_2, \ldots, T_d$ be positive integers, with $d \geq 1$

- The rectangular region, $\mathcal{R}_d = [0, T_1] \times [0, T_2] \times \cdots \times [0, T_d]$
- The r.v.'s $X_{s_1,s_2,\ldots,s_d}$, $1 \leq s_j \leq T_j$, $j \in \{1, 2, \ldots, d\}$

Let $2 \leq m_j \leq T_j$, $1 \leq j \leq d$, be positive integers

- Define for $1 \leq i_l \leq T_l - m_l + 1$, $1 \leq l \leq d$,

$$Y_{i_1,i_2,\ldots,i_d} = \sum_{s_1=i_1}^{i_1+m_1-1} \sum_{s_2=i_2}^{i_2+m_2-1} \cdots \sum_{s_d=i_d}^{i_d+m_d-1} X_{s_1,s_2,\ldots,s_d}$$

- The d-dimensional discrete scan statistic,

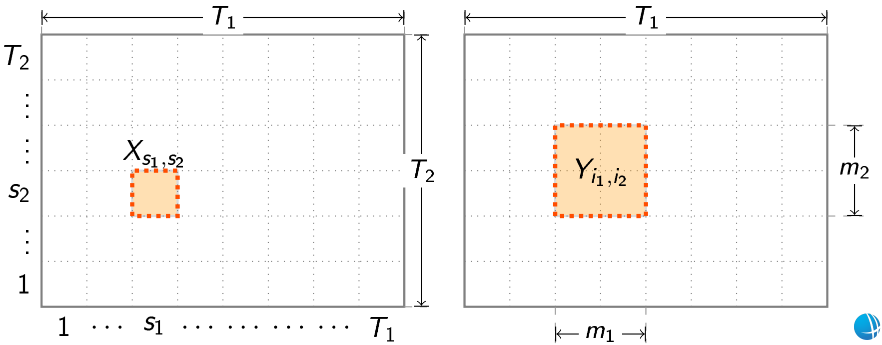
$$S_{\mathbf{m}}(\mathbf{T}) = \max_{\substack{1 \leq i_j \leq T_j - m_j + 1 \\ j \in \{1,2,\ldots,d\}}} Y_{i_1,i_2,\ldots,i_d}$$

with $\mathbf{m} = (m_1, m_2, \ldots, m_d)$ and $\mathbf{T} = (T_1, T_2, \ldots, T_d)$

Example: two dimensional scan statistics ($d = 2$)

We have for $d = 2$

$$Y_{i_1,i_2} = \sum_{s_1=i_1}^{i_1+m_1-1} \sum_{s_2=i_2}^{i_2+m_2-1} X_{s_1,s_2}, \quad S_{m_1,m_2}(T_1,T_2) = \max_{\substack{1\leq i_1\leq T_1-m_1+1 \\ 1\leq i_2\leq T_2-m_2+1}} Y_{i_1,i_2}$$



Outline

Problem

Goal

Find a good estimate for the distribution of d-dimensional discrete scan statistic

$$Q_{\mathbf{m}}(\mathbf{T}) = \mathbb{P}\left(S_{\mathbf{m}}(\mathbf{T}) \leq \tau\right)$$

The distribution of $S_{\mathbf{m}}(\mathbf{T})$ is used for testing the null hypotheses of randomness against the alternative hypothesis of clustering.

Example: Bernoulli model

H_0: The r.v.'s $X_{s_1,s_2,\ldots,s_d}$ are i.i.d. $\mathcal{B}(p)$

H_1: There exists $\mathcal{R}(i_1, i_2, \ldots, i_d) = [i_1 - 1, i_1 + m_1 - 1] \times \cdots \times [i_d - 1, i_d + m_d - 1] \subset \mathcal{R}_d$ where the r.v.'s $X_{s_1,s_2,\ldots,s_d} \sim \mathcal{B}(p')$, $p' > p$ and $X_{s_1,s_2,\ldots,s_d} \sim \mathcal{B}(p)$ outside $\mathcal{R}(i_1, i_2, \ldots, i_d)$

Outline

Exact methods (Bernoulli case)

There are three main approaches used for investigating the exact distribution of the one dimensional discrete scan statistics over a sequence of binary trials:

- Combinatorial method: [Naus, 1974], [Naus, 1982]
- Finite Markov chain imbedding technique: [Fu, 2001], [Balakrishnan and Koutras, 2002], [Fu and Lou, 2003], [Wu, 2013], [Fu et al., 2012] etc.
- Conditional generating function method: [Ebneshahrashoob and Sobel, 1990], [Ebneshahrashoob et al., 2005], [Shinde and Kotwal, 2008] etc.

Product Type Approximation

Considering that $T_1 = L_1 m_1$, [Naus, 1982] gave the following approximation

$$\mathbb{P}\left(S_{m_1}(T_1) \leq \tau\right) \approx Q(2m_1)\left[\frac{Q(3m_1)}{Q(2m_1)}\right]^{\frac{T_1}{m_1}-2},$$

where $Q(2m_1) = \mathbb{P}\left(S_{m_1}(2m_1) \leq \tau\right)$ and $Q(3m_1) = \mathbb{P}\left(S_{m_1}(3m_1) \leq \tau\right)$.

$$\begin{aligned}
Q(2m_1) &= F^2(\tau; m_1, p) - \tau b(\tau+1; m_1, p)F(\tau-1; m_1, p) + m_1 p b(\tau+1; m_1, p)F(\tau-2, m_1-1),\\
Q(3m_1) &= F^3(\tau; m_1, p) - A_1 + A_2 + A_3 - A_4,\\
A_1 &= 2b(\tau+1; m_1, p)F(\tau; m_1, p)[\tau F(\tau-1; m_1, p) - m_1 p F(\tau-2; m_1-1, p)],\\
A_2 &= 0.5 b^2(\tau+1; m_1, p)\left[\tau(\tau-1)F(\tau-2; m_1, p) - 2(\tau-1)m_1 F(\tau-3; m_1-1, p)\right.\\
&\quad \left.+ m_1(m_1-1)p^2 F(\tau-4; m_1-2, p)\right],\\
A_3 &= \sum_{r=1}^{\tau} b(2(\tau+1)-r; m_1, p)F^2(r-1; m_1, p),\\
A_4 &= \sum_{r=2}^{\tau} b(2(\tau+1)-r; m_1, p)b(r+1; m_1, p)[rF(r-1; m_1, p) - m_1 p F(r-2; m_1-1, p)].
\end{aligned}$$

Product Type Approximation for Binomial and Poisson

If $X_i \sim Bin(n,p)$ or $X_i \sim Pois(\lambda)$, we have the approximation

$$\mathbb{P}\left(S_{m_1}(T_1) \leq \tau\right) \approx Q(2m_1)\left[\frac{Q(3m_1)}{Q(2m_1)}\right]^{\frac{T_1}{m_1}-2}, \quad T_1 \geq 3m_1$$
$$\approx Q(2m_1-1)\left[\frac{Q(2m_1)}{Q(2m_1-1)}\right]^{T_1-2m_1+1}, \quad T_1 \geq 2m_1$$

where $Q(2m_1-1)$, $Q(2m_1)$ and $Q(3m_1)$ are computed by [Karwe and Naus, 1997] recurrence.

▸ Karwe Naus algorithm for $Q(2m_1-1)$ and $Q(2m_1)$

Bounds

[Glaz and Naus, 1991] developed a variety of tight bounds:

- Lower Bounds

$$\begin{aligned}\mathbb{P}\left(S_{m_1}(T_1)\leq\tau\right) &\leq \frac{Q(2m_1)}{\left[1+\frac{Q(2m_1-1)-Q(2m_1)}{Q(2m_1-1)Q(2m_1)}\right]^{T_1-2m_1}}, \quad T_1\geq 2m_1 \\ &\leq \frac{Q(3m_1)}{\left[1+\frac{Q(2m_1-1)-Q(2m_1)}{Q(3m_1-1)}\right]^{T_1-3m_1}}, \quad T_1\geq 3m_1\end{aligned}$$

- Upper Bounds

$$\begin{aligned}\mathbb{P}\left(S_{m_1}(T_1)\leq\tau\right) &\leq Q(2m_1)\left[1-Q(2m_1-1)+Q(2m_1)\right]^{T_1-2m_1}, \quad T_1\geq 2m_1 \\ &\leq Q(3m_1)\left[1-Q(2m_1-1)+Q(2m_1)\right]^{T_1-3m_1}, \quad T_1\geq 3m_1\end{aligned}$$

The values $Q(2m_1-1), Q(2m_1), Q(3m_1-1), Q(3m_1)$ are computed using [Karwe and Naus, 1997] algorithm.

Outline

Product Type Approximation Bernoulli Case

[Boutsikas and Koutras, 2000] using Markov Chain Imbedding approach proposed the approximation

$$\mathbb{P}\left(S_{m_1,m_2}(T_1,T_2)\leq\tau\right)\approx\frac{Q(m_1,m_2)^{(T_1-m_1-1)(T_2-m_2-1)}Q(m_1+1,m_2+1)^{(T_1-m_1)(T_2-m_2)}}{Q(m_1,m_2+1)^{(T_1-m_1-1)(T_2-m_2)}Q(m_1+1,m_2)^{(T_1-m_1)(T_2-m_2-1)}}$$

Where,

$$\begin{aligned}
Q(m_1,m_2) &= F\left(\tau;m_1m_2,p\right)\\
Q(m_1+1,m_2) &= \sum_{s=0}^{\tau}F^2\left(\tau-s;m_2,p\right)b\left(s;(m_1-1)m_2,p\right)\\
Q(m_1+1,m_2+1) &= \sum_{s_1,s_2=0}^{\tau}\sum_{t_1,t_2=0}^{\tau}\sum_{i_1,i_2,i_3,i_4=0}^{1}b(s_1;m_1-1,p)b(s_2;m_1-1,p)b(t_1;m_2-1,p)\times\\
&\quad b(t_2;m_2-1,p)p^{\sum i_j}(1-p)^{4-\sum i_j}F(u;(m_1-1)(m_2-1),p)\\
u &= \min\left\{\tau-s_1-t_1-i_1,\tau-s_2-t_1-i_2,\tau-s_1-t_2-i_3,\tau-s_2-t_2-i_4\right\}\\
b(s;n,p) &= \binom{n}{s}p^s(1-p)^{n-s}\\
F(s;n,p) &= \sum_{i=0}^{s}b(i;n,p)
\end{aligned}$$

Université Lille1 Sciences et Technologies

Bounds for the Bernoulli Case

The following bounds were established by [Boutsikas and Koutras, 2003]

- Lower Bound

$$LB = (1 - Q_1)^{(T_1 - m_1)(T_2 - m_2)} (1 - Q_2)^{T_1 - m_1} (1 - Q_3)^{T_2 - m_2} (1 - Q_4)$$

- Upper Bound

$$\begin{aligned} UB = & (1 - Q_1)\left(1 - q^{(m_1-1)(3m_2-2)+(2m_1-1)(m_2-1)} Q_1\right)^{(T_1-m_1-1)(T_2-m_2-1)} \left(1 - q^{m_1(m_2-1)} Q_1\right)^{T_2-m_2-1} \\ & \times \left(1 - q^{(m_1-1)(2m_2-1)+(m_1-1)(m_2-1)} Q_1\right)^{T_1-m_1-1} \left(1 - q^{(m_1-1)(2m_2-1)+m_1(m_2-1)} Q_2\right)^{T_1-m_1} \\ & \times \left(1 - q^{(m_1-1)(3m_2-2)+m_1(m_2-1)+(m_1-1)(m_2-1)} Q_3\right)^{T_2-m_2} \left(1 - q^{(m_1-1)(2m_2-1)+m_1(m_2-1)} Q_4\right). \end{aligned}$$

Where $X_{ij} \sim B(p)$, $q = 1 - p$ and

$$Q_1 = F^c_{\tau+1,m_1m_2} - q^{m_2} F^c_{\tau+1,(m_1-1)m_2} - q^{m_1} F^c_{\tau+1,m_1(m_2-1)} + q^{m_1+m_2-1} F^c_{\tau+1,(m_1-1)(m_2-1)},$$

$$Q_2 = F^c_{\tau+1,m_1m_2} - q^{m_2} F^c_{\tau+1,(m_1-1)m_2}, \quad Q_3 = F^c_{\tau+1,m_1m_2} - q^{m_1} F^c_{\tau+1,m_1(m_2-1)}, \quad Q_4 = F^c_{\tau+1,m_1m_2}$$

$$F^c_{i,m} = 1 - F(i-1; m, p).$$

Product Type Approximation Binomial and Poisson

For $X_{ij} \sim Bin(n, p)$ or $X_{ij} \sim Pois(\lambda)$, [Chen and Glaz, 2009] proposed the product type approximation

$$\mathbb{P}\left(S_{m_1,m_2}(T_1, T_2) \leq \tau\right) \approx \frac{Q(m_1+1,m_2+1)^{(T_1-m_1)(T_2-m_2)}}{Q(m_1+1,m_2)^{(T_1-m_1)(T_2-m_2-1)}} \times \frac{Q(m_1,2m_2-1)^{(T_1-m_1-1)(T_2-2m_2)}}{Q(m_1,2m_2)^{(T_1-m_1-1)(T_2-2m_2+1)}}$$

Where,

$$Q(m_1, 2m_2 - 1) = \mathbb{P}\left(S_{m_1,m_2}(m_1, 2m_2 - 1) \leq \tau\right)$$
$$Q(m_1, 2m_2) = \mathbb{P}\left(S_{m_1,m_2}(m_1, 2m_2) \leq \tau\right)$$

To compute the unknown variables we use

- $Q(m_1, 2m_2 - 1)$ and $Q(m_1, 2m_2)$ - adaptation of [Karwe and Naus, 1997] algorithm
- $Q(m_1 + 1, m_2)$ and $Q(m_1 + 1, m_2 + 1)$ - conditioning

Formulas for $Q(m_1 + 1, m_2)$ and $Q(m_1 + 1, m_2 + 1)$

Lower Bound for Binomial and Poisson

[Chen and Glaz, 1996] gave a lower bound applying Hoover Bonferroni type inequality of order $r \geq 3$,

$$\mathbb{P}\left(S_{m_1,m_2}(T_1,T_2) \geq \tau\right) = \mathbb{P}\left(\bigcup_{i_1=1}^{T_1-m_1+1} \bigcup_{i_2=1}^{T_2-m_2+1} A_{i_1,i_2}\right) \leq \sum_{i_1=1}^{T_1-m_1+1} \sum_{i_2=1}^{T_2-m_2+1} \mathbb{P}\left(A_{i_1,i_2}\right)$$
$$- \sum_{i_1=1}^{T_1-m_1+1} \sum_{i_2=1}^{T_2-m_2} \mathbb{P}\left(A_{i_1,i_2} \cap A_{i_1,i_2+1}\right) - \sum_{i_1=1}^{T_1-m_1} \mathbb{P}\left(A_{i_1,1} \cap A_{i_1+1,1}\right)$$
$$- \sum_{i_1=1}^{T_1-m_1+1} \sum_{l=2}^{r-1} \sum_{i_2=1}^{T_2-m_2+1-l} \mathbb{P}\left(A_{i_1,i_2} \cap A^c_{i_1,i_2+1} \cdots A^c_{i_1,i_2+l-1} \cap A_{i_1,i_2+l}\right)$$

Where $A_{i_1,i_2} = \{Y_{i_1,i_2} \geq \tau\}$ and for $r = 4$,

- Lower Bound

$$\mathbb{P}\left(S_{m_1,m_2}(T_1,T_2) \leq \tau\right) \geq (T_1 - m_1)\left(Q(m_1+1,m_2) - 2Q(m_1,m_2)\right) - (T_1 - m_1 + 1)(T_2 - m_2 - 3)$$
$$\times Q(m_1, m_2+2) + (T_1 - m_1 + 1)(T_2 - m_2 - 2)Q(m_1, m_2+3).$$

- $Q(m_1+1, m_2)$, $Q(m_1, m_2)$, $Q(m_1, m_2+2)$, $Q(m_1, m_2+3)$ - [Karwe and Naus, 1997] algorithm (variant)

Upper Bound for Binomial and Poisson

For the upper bound we adapt the inequality of [Kuai et al., 2000] to the two dimensional framework:

$$\mathbb{P}\left(S_{m_1,m_2}(T_1,T_2)\le\tau\right)=1-\mathbb{P}\left(\bigcup_{i_1=1}^{T_1-m_1+1}\bigcup_{i_2=1}^{T_2-m_2+1}A_{i_1,i_2}\right)$$

$$\le 1-\sum_{i_1=1}^{T_1-m_1+1}\sum_{i_2=1}^{T_2-m_2+1}\left[\frac{\theta_{i_1,i_2}\mathbb{P}(A_{i_1,i_2})^2}{\Sigma(i_1,i_2)+(1-\theta_{i_1,i_2})\mathbb{P}(A_{i_1,i_2})}+\frac{(1-\theta_{i_1,i_2})\mathbb{P}(A_{i_1,i_2})^2}{\Sigma(i_1,i_2)-\theta_{i_1,i_2}\mathbb{P}(A_{i_1,i_2})}\right]$$

where

$$\Sigma(i_1,i_2)=\sum_{j_1=1}^{T_1-m_1+1}\sum_{j_2=1}^{T_2-m_2+1}\mathbb{P}\left(A_{i_1,i_2}\cap A_{j_1,j_2}\right)\text{ and }\theta_{i_1,i_2}=\frac{\Sigma(i_1,i_2)}{\mathbb{P}(A_{i_1,i_2})}-\left\lfloor\frac{\Sigma(i_1,i_2)}{\mathbb{P}(A_{i_1,i_2})}\right\rfloor.$$

We have

$$\mathbb{P}\left(A_{i_1,i_2}\cap A_{j_1,j_2}\right)=\begin{cases}[1-Q(m_1,m_2)]^2, & \text{if } |i_1-j_1|\ge m_1 \text{ or } |i_2-j_2|\ge m_2,\\ 1-2Q(m_1,m_2)+\mathbb{P}\left(Y_{i_1,i_2}\le\tau,Y_{j_1,j_2}\le\tau\right), & \text{otherwise}\end{cases}$$

$$\mathbb{P}\left(Y_{i_1,i_2}\le\tau,Y_{j_1,j_2}\le\tau\right)=\sum_{k=0}^{\tau}\mathbb{P}(Z=k)\mathbb{P}(Y_{i_1,i_2}-Z\le\tau-k)^2,$$

$$Z=\sum_{s=i_1\vee j_1}^{(i_1+m_1-1)\wedge(j_1+m_1-1)}\sum_{t=i_2\vee j_2}^{(i_2+m_2-1)\wedge(j_2+m_2-1)}X_{s,t}.$$

Outline

Product Type Approximation

Glaz et al. proposed in [Guerriero et al., 2010] the product type approximation

$$\mathbb{P}\left(S_{m_1,m_2,m_3}(T_1, T_2, T_3) \leq \tau\right) \approx$$

$$\frac{Q(m_1+1, m_2+1, m_3+1)^{(T_1-m_1)(T_2-m_2)(T_3-m_3)} Q(m_1+1, m_2, m_3)^{(T_1-m_1)(T_2-m_2-1)(T_3-m_3-1)}}{Q(m_1, m_2, m_3)^{(T_1-m_1-1)(T_2-m_2-1)(T_3-m_3-1)} Q(m_1+1, m_2+1, m_3)^{(T_1-m_1)(T_2-m_2)(T_3-m_3-1)}} \times$$

$$\frac{Q(m_1, m_2+1, m_3)^{(T_1-m-1-1)(T_2-m_2)(T_3-m_3-1)} Q(m_1, m_2, m_3+1)^{(T_1-m_1-1)(T_2-m-2-1)(T_3-m_3-1)}}{Q(m_1+1, m_2, m_3+1)^{(T_1-m-1)(T-2-m_2-1)(T-3-m-3)} Q(m_1, m_2+1, m_3+1)^{(T_1-m_1-1)(T_2-m_2)(T_3-m-3)}}$$

Where,

$$Q(N_1, N_2, N_3) = \mathbb{P}\left(S_{m_1,m_2,m_3}(N_1, N_2, N_3) \leq \tau\right)$$

- The approximation also works for binomial and Poisson distribution
- Three Poisson Type Approximation

Lower and upper bounds for the distribution of three dimensional scan statistics were proposed by [Akiba and Yamamoto, 2004].

Université Lille1 Sciences et Technologies

Outline

Haiman Type Approximation: Key Idea

Haiman proposed in [Haiman, 2000] a different approach

Main Observation

The scan statistic r.v. can be viewed as a maximum of a sequence of 1-dependent stationary r.v..

- The idea:
  - discrete and continuous one dimensional scan statistic: [Haiman, 2000], [Haiman, 2007]
  - discrete and continuous two dimensional scan statistic: [Haiman and Preda, 2002], [Haiman and Preda, 2006]
  - discrete three dimensional scan statistic: [Amărioarei and Preda, 2013]

Expressing $S_{\mathbf{m}}(\mathbf{T})$ as a maximum of a 1-dependent sequence

Let $L_j = \frac{T_j}{m_j - 1}$, $j \in \{1, 2, \ldots, d\}$, be positive integers

- Define for each $k_1 \in \{1, 2, \ldots, L_1 - 1\}$ the random variables

$$Z_{k_1} = \max_{\substack{(k_1-1)(m_1-1)+1 \leq i_1 \leq k_1(m_1-1) \\ 1 \leq i_j \leq (L_j-1)(m_j-1) \\ j \in \{2,\ldots,d\}}} Y_{i_1, i_2, \ldots, i_d}$$

- $(Z_j)_j$ is 1-dependent and stationary

Example: $d = 1$

$$X_1, X_2, \ldots, X_{m_1-1}, X_{m_1}, \ldots, X_{2(m_1-1)}, X_{2m_1-1}, \ldots, X_{3(m_1-1)}, X_{3m_1-2}, \ldots, X_{4(m_1-1)}$$

Z_2

Z_1

Z_3

- Observe

$$S_{\mathbf{m}}(\mathbf{T}) = \max_{1 \leq k_1 \leq L_1 - 1} Z_{k_1}$$

The main tool

Let $(Z_j)_{j\geq 1}$ be a strictly stationary 1-dependent sequence of r.v.'s and let $q_m = q_m(x) = \mathbb{P}(\max(Z_1, \dots, Z_m) \leq x)$, with $x < \sup\{u | \mathbb{P}(Z_1 \leq u) < 1\}$.

Theorem [Amărioarei, 2012]

For x such that $\mathbb{P}(Z_1 > x) = 1 - q_1 < 0.1$ and $m > 3$ we have

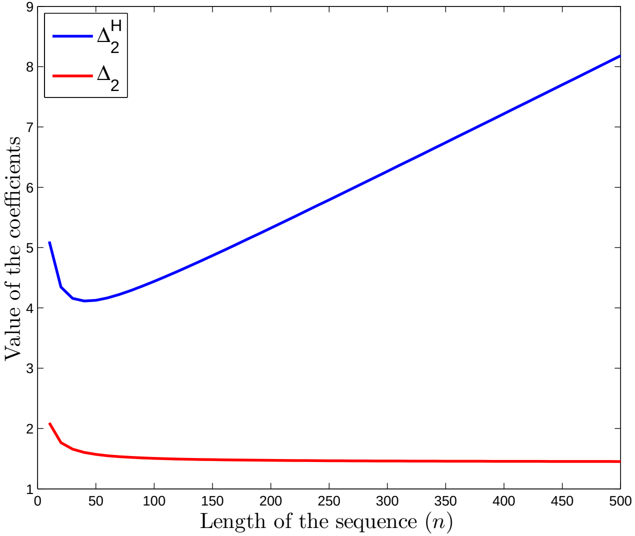
$$\left| q_m - \frac{2q_1 - q_2}{[1 + q_1 - q_2 + 2(q_1 - q_2)^2]^m} \right| \leq \Delta_2 (1 - q_1)^2$$

- $\Delta_2 = mF(q_1, m) = m\left[1 + \frac{3}{m} + K(\alpha)(1 - q_1) + \frac{\Gamma(\alpha)(1-q_1)}{m}\right]$.

- Increased range of applicability
- Sharp bounds values (ex. $\alpha = 0.025$: $561 \rightarrow 145$ and $88 \rightarrow 17.5$)

Selected values for $K(\alpha)$ and $\Gamma(\alpha)$

Difference between the results: $1 - q_1 = 0.025$



Approximation process

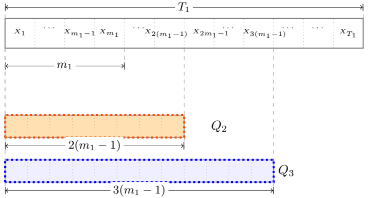
Define for $t_1 \in \{2,3\}$,

$$Q_{t_1} = Q_{t_1}(\tau) = \mathbb{P}\left(\bigcap_{k_1=1}^{t_1-1}\{Z_{k_1} \leq \tau\}\right) = \mathbb{P}\left(\max_{\substack{1\leq i_1\leq (t_1-1)(m_1-1)\\ 1\leq i_j\leq (L_j-1)(m_j-1)\\ j\in\{2,\dots,d\}}} Y_{i_1,i_2,\dots,i_d} \leq \tau\right)$$

If $1 - Q_2 \leq 0.1$ then

$$\left|Q_{\mathbf{m}}(\mathbf{T}) - \frac{2Q_2-Q_3}{[1+Q_2-Q_3+2(Q_2-Q_3)^2]^{L_1-1}}\right| \leq (L_1-1)F(Q_2, L_1-1)(1-Q_2)^2$$

Example: $d = 1$



- The approximation

$$\mathbb{P}(S_{m_1}(T_1) \leq \tau) \approx H(Q_2, Q_3, L_1)$$

$$H(x,y,m) = \frac{2x-y}{[1+x-y+2(x-y)^2]^{m-1}}$$

- Approximation error, about

$$(L_1-1)F(Q_2, L_1-1)(1-Q_2)^2$$

Approximation process 2

The approximation of $S_{\mathbf{m}}(\mathbf{T})$ is an iterative process. The s step, $1 \leq s \leq d$, becomes:

- Let

$$Q_{t_1,t_2,\ldots,t_s} = Q_{t_1,t_2,\ldots,t_s}(\tau) = \mathbb{P}\left(\max_{\substack{1\leq i_l\leq (t_l-1)(m_l-1)\\ l\in\{1,\ldots,s\}\\ 1\leq i_j\leq (L_j-1)(m_j-1)\\ j\in\{s+1,\ldots,d\}}} Y_{i_1,i_2,\ldots,i_d} \leq \tau\right)$$

- Define for $t_l \in \{2,3\}$, $l \in \{1,\ldots,s-1\}$ and $k_s \in \{1,2,\ldots,L_s-1\}$

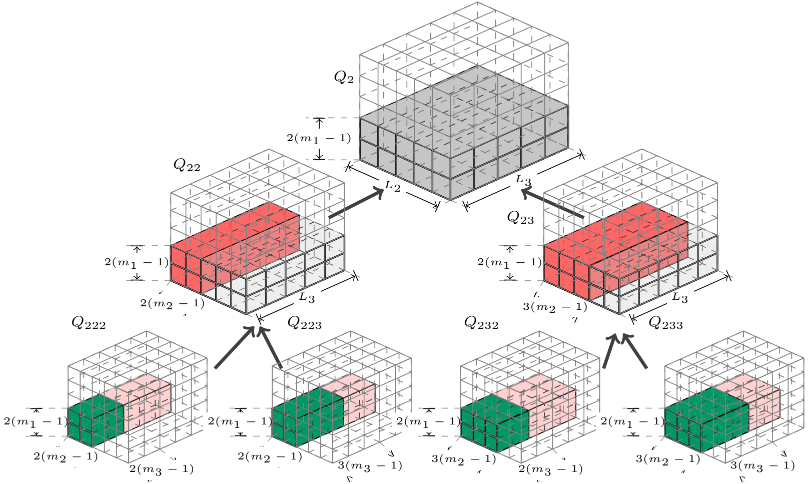
$$Z_{k_s}^{(t_1,t_2,\ldots,t_{s-1})} = \max_{\substack{1\leq i_l\leq (t_l-1)(m_l-1)\\ l\in\{1,2,\ldots,s-1\}\\ (k_s-1)(m_s-1)+1\leq i_s\leq k_s(m_s-1)\\ 1\leq i_j\leq (L_j-1)(m_j-1)\\ j\in\{s+1,\ldots,d\}}} Y_{i_1,i_2,\ldots,i_d}$$

- $\left\{Z_1^{(t_1,t_2,\ldots,t_{s-1})},\ldots,Z_{L_s-1}^{(t_1,t_2,\ldots,t_{s-1})}\right\}$ forms a 1-dependent stationary sequence
- The approximation

$$\left|Q_{t_1,\ldots,t_{s-1}} - H\left(Q_{t_1,\ldots,t_{s-1},2}, Q_{t_1,\ldots,t_{s-1},3}, L_s\right)\right| \leq (L_s-1)F(Q_{t_1,\ldots,t_{s-1},2}, L_s-1)(1-Q_{t_1,\ldots,t_{s-1},2})^2$$

Error bounds

Illustration of the approximation process: $d = 3$



Outline

Comparison between methods: $d = 1$

Table 1 : $n = 1, p = 0.005, m_1 = 10, T_1 = 1000, It_{App} = 10^4$

τ	Exact	Glaz and Naus Product type	Haiman Approximation	Approximation Error	Lower Bound	Upper Bound
1	0.810209	0.810216	0.810404	0.001111	0.809903	0.810439
2	0.995764	0.995764	0.995764	3×10^{-7}	0.995764	0.995764
3	0.999950	0.999950	0.999950	4×10^{-11}	0.999950	0.999950

Table 2 : $n = 5, p = 0.05, m_1 = 25, T_1 = 500, It_{App} = 10^4, It_{Sim} = 10^3$

τ	$\hat{\mathbb{P}}(S \leq \tau)$	Glaz and Naus Product type	Haiman Approximation	Total Error	Lower Bound	Upper Bound
13	0.712750	0.705787	0.714699	0.039308	0.697431	0.706948
14	0.867498	0.862184	0.865029	0.012502	0.859543	0.862407
15	0.946912	0.943329	0.946177	0.004169	0.942552	0.943362
16	0.980230	0.978959	0.979822	0.001354	0.978733	0.978963
17	0.993486	0.992821	0.993134	0.000433	0.992756	0.992822
18	0.997802	0.997726	0.997849	0.000127	0.997708	0.997726
19	0.999362	0.999327	0.999358	3×10^{-5}	0.999322	0.999327
20	0.999819	0.999813	0.999825	9×10^{-6}	0.999812	0.999813
21	0.999954	0.999951	0.999953	2×10^{-6}	0.999951	0.999951

Comparison between methods: $d = 2$

Table 3 : $n = 1, p = 0.005, m_1 = m_2 = 6, T_1 = T_2 = 30, It_{App} = 10^3, It_{Sim} = 10^3$

τ	$\hat{\mathbb{P}}(S \leq \tau)$	Glaz and Naus Product type	Haiman Approximation	Total Error(App+Sim)	Lower Bound	Upper Bound
2	0.915903	0.914013	0.920211	0.041483	0.901935	0.945623
3	0.994292	0.994395	0.994578	0.000803	0.993785	0.996638
4	0.999747	0.999757	0.999760	2×10^{-5}	0.999737	0.999858
5	0.999992	0.999992	0.999992	7×10^{-7}	0.999992	0.999995

Table 4 : $n = 5, p = 0.002, m_1 = 5, m_2 = 10, T_1 = 50, T_2 = 80, It_{App} = 10^4, It_{Sim} = 10^3$

τ	$\hat{\mathbb{P}}(S \leq \tau)$	Glaz and Naus Product type	Haiman Approximation	Total Error(App+Sim)	Lower Bound	Upper Bound
4	0.894654	0.873256	0.893724	0.037136	0.803422	0.944318
5	0.988003	0.986249	0.988144	0.002125	0.981418	0.993451
6	0.998963	0.998847	0.998963	0.000152	0.998543	0.999401
7	0.999926	0.999919	0.999925	9×10^{-6}	0.999903	0.999955
8	0.999995	0.999995	0.999995	5×10^{-7}	0.999994	0.999997

Université Lille1 Sciences et Technologies

Comparison between methods: $d = 3$

Table 5 : $n = 1, p = 0.00005, m_1 = m_2 = m_3 = 5, T_1 = T_2 = T_3 = 60, It_{App} = 10^5$

τ	$\hat{\mathbb{P}}(S \leq \tau)$	Glaz et al. Product type	Haiman Approximation	Approximation Error	Simulation Error	Total Error
1	0.841806	0.841424	0.851076	0.011849	0.064889	0.076738
2	0.999119	0.999142	0.999192	0.000000	0.000170	0.000170
3	0.999997	0.999998	0.999997	0.000000	3×10^{-7}	3×10^{-7}

Table 6 : $n = 1, p = 0.0001, m_1 = m_2 = m_3 = 5, T_1 = T_2 = T_3 = 60, It_{App} = 10^5$

τ	$\hat{\mathbb{P}}(S \leq \tau)$	Glaz et al. Product type	Haiman Approximation	Approximation Error	Simulation Error	Total Error
2	0.993294	0.993241	0.993192	0.000010	0.001367	0.001377
3	0.999963	0.999964	0.999963	0.000000	0.000005	0.000005
4	0.999999	0.999999	0.999999	0.000000	2×10^{-9}	2×10^{-9}

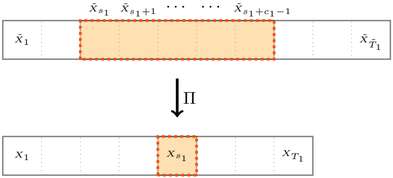
Outline

Introducing the model $d = 1$

Let $m_1, c_1, \tilde{T}_1$ be positive integers such that $c_1 \geq 1$ and $2 \leq m_1 \leq T_1$

- Consider $\left(\tilde{X}_i\right)_{1 \leq i \leq \tilde{T}_1}$ be a sequence of i.i.d. r.v's
- Take $\Pi : \mathbb{R}^{c_1} \to \mathbb{R}$ to be a measurable function
- Define $T_1 = \tilde{T}_1 - c_1 + 1$ and for $1 \leq s_1 \leq T_1$

$$X_{s_1} = \Pi\left(\tilde{X}_{s_1}, \dots, \tilde{X}_{s_1+c_1-1}\right)$$



- The sequence $(X_{s_1})_{1 \leq s_1 \leq T_1}$ is $(c_1 - 1)$-dependent

Approximation and bounds

Assume that $L_1 = \frac{\tilde{T}_1}{m_1+c_1-2}$ is a positive integer

- Define for $k_1 \in \{1, 2, \dots, L_1 - 1\}$

$$Z_{k_1} = \max_{(k_1-1)(m_1+c_1-2)+1 \leq i_1 \leq k_1(m_1+c_1-2)} Y_{i_1}$$

- $(Z_j)_j$ is 1-dependent and stationary
- We have the approximation

$$\mathbb{P}(S_{m_1}(T_1) \leq \tau) \approx \frac{2Q_2 - Q_3}{[1+Q_2-Q_3+2(Q_2-Q_3)^2]^{L_1-1}}$$

- Error bounds

$$E_{app}(1) = (L_1-1)F(Q_2, L_1-1)(1-Q_2)^2,$$

$$E_{sapp}(1) = (L_1-1)F\left(\hat{Q}_2, L_1-1\right)\left(1-\hat{Q}_2+\beta_2\right)^2,$$

$$E_{sf}(1) = (L_1-1)(\beta_2+\beta_3)$$

Outline

Longest increasing run

Let $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_{\tilde{T}_1}$ be a sequence of i.i.d. r.v.'s with the common distribution G.
A subsequence $\left(\tilde{X}_k, \ldots, \tilde{X}_{k+l-1}\right)$ forms an *increasing run* (or ascending run) of length $l \geq 1$, starting at position $k \geq 1$, if it verifies the following relation

$$\tilde{X}_{k-1} > \tilde{X}_k < \tilde{X}_{k+1} < \cdots < \tilde{X}_{k+l-1} > \tilde{X}_{k+l}$$

Denote the length of the longest increasing run among the first $\tilde{T}_1$ random variables by $M_{\tilde{T}_1}$.
The asymptotic distribution was studied

- G continuous distribution: [Pittel, 1981], [Révész, 1983], [Grill, 1987], [Novak, 1992], etc.
- G discrete distribution: [Csaki and Foldes, 1996], [Grabner et al., 2003], [Eryilmaz, 2006], etc.

Longest increasing run 2

Let $c_1 = 2$, $T_1 = \tilde{T}_1 - 1$ and define $\Pi : \mathbb{R}^2 \to \mathbb{R}$ by

$$\Pi(x, y) = \begin{cases} 1, \text{ if } x < y \\ 0, \text{ otherwise} \end{cases}$$

Thus our block-factor model becomes

$$X_{s_1} = \mathbf{1}_{\tilde{X}_{s_1} < \tilde{X}_{s_1+1}}$$

Let L_{T_1} be the distribution of the length of the longest run of ones, among the first T_1 observations.

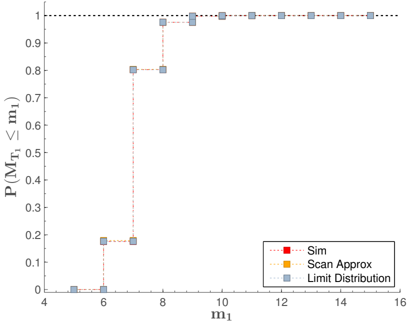
$$\mathbb{P}\left(M_{\tilde{T}_1} \leq m_1\right) = \mathbb{P}\left(L_{T_1} < m_1\right) = \mathbb{P}\left(S_{m_1}(T_1) < m_1\right), \text{ for } m_1 \geq 1$$

Longest increasing run: numerical results

For $\tilde{X}_{s_1} \sim \mathcal{U}([0,1])$, [Novak, 1992] showed that

$$\max_{1 \le m_1 \le T_1} \left| \mathbb{P}\left(L_{T_1} < m_1\right) - e^{-T_1 \frac{m_1+1}{(m_1+2)!}} \right| = \mathcal{O}\left(\frac{\ln T_1}{T_1}\right)$$

m_1	Sim	AppH	$E_{total}(1)$	LimApp
5	0.00000700	0.00000733	0.14860299	0.00000676
6	0.17567262	0.17937645	0.01089628	0.17620431
7	0.80257424	0.80362353	0.00110990	0.80215088
8	0.97548510	0.97566460	0.00011579	0.97550345
9	0.99749821	0.99751049	0.00001114	0.99749792
10	0.99977074	0.99977183	0.00000098	0.99977038
11	0.99998075	0.99998083	0.00000008	0.99998073
12	0.99999851	0.99999851	0.00000001	0.99999851
13	0.99999989	0.99999989	0.00000000	0.99999989
14	0.99999999	0.99999999	0.00000000	0.99999999
15	1.00000000	1.00000000	0.00000000	1.00000000



Longest increasing run: numerical results

For $\tilde{X}_{s_1} \sim Geom(p)$, [Louchard and Prodinger, 2003] showed that

$$\mathbb{P}\left(M_{T_1} \leq m_1\right) \sim \exp\left(-\exp\eta\right),$$
$$\eta = \frac{m_1(m_1+1)}{2}\log\frac{1}{1-p} + m_1\log\frac{1}{p} - \log T_1 - \log p + \log D(m_1),$$
$$D(m_1) = \prod_{k=1}^{m_1}\left[1-(1-p)^k\right]\left[1-(1-p)^{m_1+2}\right]$$

m_1	Sim	AppH	$E_{total}(1)$	LimApp
6	0.56445934	0.56997462	0.00255592	0.56810748
7	0.95295406	0.95325180	0.00018554	0.95294598
8	0.99658057	0.99659071	0.00001214	0.99657969
9	0.99979460	0.99979550	0.00000068	0.99979435
10	0.99998950	0.99998950	0.00000003	0.99998947

We used $T_1 = 10000$, $p = 0.1$ and $Iter = 10^5$.

Université Lille1 Sciences et Technologies

Akiba, T. and Yamamoto, H. (2004).
Upper and lower bounds for 3-dimensional r-within n consecutive -(r_1, r_2, r_3)-out-of -(n_1, n_2, n_3): F-system.
In *Advanced Reliability Modelling*.

Amărioarei, A. (2012).
Approximation for the distribution of extremes of one dependent stationary sequences of random variables.
arXiv:1211.5456v1, submitted.

Amărioarei, A. and Preda, C. (2013).
Approximation for the distribution of three-dimensional discrete scan statistic.
Methodol Comput Appl Probab, DOI 10.1007/s11009-013-9382-3.

Balakrishnan, N. and Koutras, M. V. (2002).
Runs and scans with applications.
Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York.

Boutsikas, M. V. and Koutras, M. V. (2000).
Reliability approximation for Markov chain imbeddable systems.
Methodol. Comput. Appl. Probab., 2:393–411.

Boutsikas, M. V. and Koutras, M. V. (2003).
Bounds for the distribution of two-dimensional binary scan statistics.
Probab. Eng. Inform. Sci., 17:509–525.

Chen, J. and Glaz, J. (1996).
Two-dimensional discrete scan statistics.
Statist. Probab. Lett., 31:59–68.

Chen, J. and Glaz, J. (2009).
Scan statistics, chapter 5, Approximations for two-dimensional variable window scan statistics., pages 109–128.
Birkhäuser Boston, Inc., Boston.

Csaki, E. and Foldes, A. (1996).
On the length of theh longest monnotone block.
Studio Scientiarum Mathematicarum Hungarica, 31:35–46.

Ebneshahrashoob, M., Gao, T., and Wu, M. (2005).
An efficient algorithm for exact distribution of discrete scan statistics.
Methodol. Comput. Appl. Probab., 7:1423–1436.

Ebneshahrashoob, M. and Sobel, M. (1990).
Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas.
Statist. Probab. Lett., 9:5–11.

Eryilmaz, S. (2006).
A note on runs of geometrically distributed random variables.
Discrete Mathematics, 306:1765–1770.

Fu, J. (2001).
Distribution of the scan statistic for a sequence of bistate trials.
J. Appl. Probab., 38:908–916.

Fu, J., Wu, T.-L., and Lou, W. (2012).
Continuous, discrete, and conditional scan statistics.
J. Appl. Prob., 49:199–209.

Fu, J. C. and Lou, W. (2003).
Distribution theory of runs and patterns and its applications. A finite Markov chain imbedding approach.
World Scientific Publishing Co., Inc., River Edge, NJ.

Glaz, J. and Naus, J. (1991).
Tight bounds and approximations for scan statistic probabilities for discrete data.
Annals of Applied Probability, 1:306–318.

Grabner, P., Knopfmacher, A., and Prodinger, H. (2003).
Combinatorics of geometrically distributed random variables: run statistics.
Theoret. Comput. Sci., 297:261–270.

Grill, K. (1987).
Erdos-Révész type bounds for the length of the longest run from a stationary mixing sequence.
Probab. Theory Relat. Fields, 75:169–179.

Guerriero, M., Glaz, J., and Sen, R. (2010).
Approximations for a three dimensional scan statistic.
Methodol. Comput. Appl. Probab., 12:731–747.

Haiman, G. (2000).
Estimating the distributions of scan statistics with high precision.
Extremes, 3:349–361.

Haiman, G. (2007).
Estimating the distribution of one-dimensional discrete scan statistics viewed as extremes of 1-dependent stationary sequences.
J. Statist. Plann. Inference, 137:821–828.

Haiman, G. and Preda, C. (2002).
A new method for estimating the distribution of scan statistics for a two-dimensional Poisson process.
Methodol. Comput. Appl. Probab., 4:393–407.

Haiman, G. and Preda, C. (2006).

Estimation for the distribution of two-dimensional discrete scan statistics.
Methodol. Comput. Appl. Probab., 8:373–381.

Karwe, V. and Naus, J. (1997).
New recursive methods for scan statistic probabilities.
Computational Statistics & Data Analysis, 17:389–402.

Kuai, H., Alajaji, F., and Takahara, G. (2000).
A lower bound on the probability of a finite union of events.
Discrete Mathematics, 215:147–158.

Louchard, G. and Prodinger, H. (2003).
Ascending runs of sequences of geometrically distributed random variables: a probabilistic analysis.
Theoret. Comput. Sci., 304:59–86.

Naus, J. (1974).
Probabilities for a generalized birthday problem.
Journal of American Statistical Association, 69:810–815.

Naus, J. (1982).
Approximations for distributions of scan statistics.
Journal of American Statistical Association, 77:177–183.

Novak, S. (1992).
Longest runs in a sequence of m-dependent random variables.
Probab. Theory Relat. Fields, 91:269–281.

Pittel, B. (1981).
Limiting behavior of a process of runs.
Ann. Probab., 9:119–129.

Révész, P. (1983).
Three problems on the llength of increasing runs.
Stochastic Process. Appl., 5:169–179.

Shinde, R. and Kotwal, K. (2008).
Distributions of scan statistics in a sequence of Markov Bernoulli trials.
International Journal of Statisitcs, LXVI:135–155.

Wu, T.-L. (2013).

On finite Markov chain imbedding technique.
Methodol Comput Appl Probab, 15:453–465.

Computation of $Q(m_1+1, m_2)$ and $Q(m_1+1, m_2+1)$

Consider $X_{ij} \sim B(n,p)$ and the notation $Y_{i_1,i_2}^{j_1,j_2} = \sum_{i_1=1}^{j_1}\sum_{i_2=1}^{j_2} X_{ij}$,

$$\mathbb{P}\left(S_{m_1,m_2}(m_1+1, m_2) \le k\right) = \sum_{y=0}^{k\wedge(m_1-1)m_2 n} \mathbb{P}^2\left(Y_{1,1}^{1,m_2} \le k-y\right)\mathbb{P}\left(Y_{2,1}^{m_1,m_2} = y\right)$$

$$\mathbb{P}\left(S_{m_1,m_2}(m_1+1, m_2+1) \le k\right) = \sum_{y_1=0}^{k\wedge(m_1-1)(m_2-1)n}\ \sum_{y_2=0}^{(k-y_1)\wedge(m_2-1)n}\ \sum_{y_3=0}^{(k-y_1)\wedge(m_2-1)n}$$

$$\sum_{y_4=0}^{[k-y_1-y_2\vee y_3]\wedge(m_1-1)n}\ \sum_{y_5=0}^{[k-y_1-y_2\vee y_3]\wedge(m_1-1)n} \mathbb{P}\left(Y_{1,1}^{1,1} \le a_1\right)$$

$$\times\, \mathbb{P}\left(Y_{1,m_2+1}^{1,m_2+1} \le a_2\right)\mathbb{P}\left(Y_{m_1+1,1}^{m_1+1,1} \le a_3\right)\mathbb{P}\left(Y_{m_1+1,m_2+1}^{m_1+1,m_2+1} \le a_4\right)$$

$$\times\, \mathbb{P}\left(Y_{2,2}^{m_1,m_2} = y_1\right)\mathbb{P}\left(Y_{1,2}^{1,m_2} = y_2\right)\mathbb{P}\left(Y_{m_1+1,2}^{m_1+1,m_2} = y_3\right)$$

$$\times\, \mathbb{P}\left(Y_{2,1}^{m_1,1} = y_4\right)\mathbb{P}\left(Y_{2,m_2+1}^{m_1,m_2+1} = y_5\right)$$

$$a_1 = k - y_1 - y_2 - y_4,\quad a_2 = k - y_1 - y_2 - y_5,$$

$$a_3 = k - y_1 - y_3 - y_4,\quad a_4 = k - y_1 - y_3 - y_5,$$

Université Lille1 Sciences et Technologies

Return

Karwe Naus recursive methods

Define

$$b^k_{2(m)}(y) = \mathbb{P}\left(S_m(2m) \leq k,\, Y_{m+1}(m) = y\right)$$
$$f(y) = \mathbb{P}(X_1 = y)$$
$$Q^k_{2m} = \mathbb{P}\left(S_m(2m) \leq k\right)$$

We have the recurrence relations

$$b^k_{2(1)}(y) = \left(\sum_{j=0}^{k} f(j)\right) f(y)$$
$$b^k_{2(m)}(y) = \sum_{\eta=0}^{y} \sum_{\nu=0}^{k-y+\eta} b^{k-\nu}_{2(m-1)}(y-\eta) f(\nu) f(\eta)$$
$$Q^k_{2m} = \sum_{y=0}^{k} b^k_{2(m)}(y)$$
$$Q^k_{2m-1} = \sum_{x=0}^{k} f(x) Q^{k-x}_{2(m-1)}$$

Return

Selected Values for $K(\alpha)$ and $\Gamma(\alpha)$

α	$K(\alpha)$	$\Gamma(\alpha)$
0.1	38.63	480.69
0.05	21.28	180.53
0.025	17.56	145.20
0.01	15.92	131.43

Table 7 : Selected values for $K(\alpha)$ and $\Gamma(\alpha)$

Return

Error bounds

Let $\gamma_{t_1,\ldots,t_d} = Q_{t_1,\ldots,t_d}$, with $t_j \in \{2,3\}$, $j \in \{1,\ldots,d\}$, and define for $2 \leq s \leq d$

$$\gamma_{t_1,\ldots,t_{s-1}} = H\left(\gamma_{t_1,\ldots,t_{s-1},2}, \gamma_{t_1,\ldots,t_{s-1},3}, L_s\right)$$

Denote with $\hat{Q}_{t_1,\ldots,t_d}$ the estimated value of $Q_{t_1,\ldots,t_d}$ and define for $2 \leq s \leq d$

$$\hat{Q}_{t_1,\ldots,t_{s-1}} = H\left(\hat{Q}_{t_1,\ldots,t_{s-1},2}, \hat{Q}_{t_1,\ldots,t_{s-1},3}, L_s\right)$$

We observe that

$$\left|Q_{\mathrm{m}}(\mathrm{T}) - H\left(\hat{Q}_2, \hat{Q}_3, L_1\right)\right| \leq \underbrace{\left|Q_{\mathrm{m}}(\mathrm{T}) - H\left(\gamma_2, \gamma_3, L_1\right)\right|}_{E_{app}(d)} + \underbrace{\left|H\left(\gamma_2, \gamma_3, L_1\right) - H\left(\hat{Q}_2, \hat{Q}_3, L_1\right)\right|}_{E_{sim}(d)}$$

Return

Error bounds: approximation and simulation errors 1

Approximation error

$$E_{app}(d) = \sum_{s=1}^{d} (L_1 - 1)\cdots(L_s - 1) \sum_{t_1,\dots,t_{s-1}\in\{2,3\}} F_{t_1,\dots,t_{s-1}} \left(1 - \gamma_{t_1,\dots,t_{s-1},2} + B_{t_1,\dots,t_{s-1},2}\right)^2,$$

where for $2 \leq s \leq d$

$$F_{t_1,\dots,t_{s-1}} = F\left(Q_{t_1,\dots,t_{s-1},2}, L_s - 1\right), \; F = F\left(Q_2, L_1 - 1\right),$$

$$B_{t_1\dots,t_{s-1}} = (L_s - 1)\left[F_{t_1,\dots,t_{s-1}} \left(1 - \gamma_{t_1,\dots,t_{s-1},2} + B_{t_1\dots,t_{s-1},2}\right)^2 + \sum_{t_s\in\{2,3\}} B_{t_1\dots,t_s}\right],$$

$$B_{t_1\dots,t_{d-1}} = (L_d - 1)F_{t_1,\dots,t_{d-1}} \left(1 - \gamma_{t_1,\dots,t_{d-1},2} + B_{t_1\dots,t_{d-1},2}\right)^2, \; B_{t_1\dots,t_d} = 0,$$

and for $s = 1$: $\sum_{t_1,t_0\in\{2,3\}} x = x$, $F_{t_1,t_0} = F$, $\gamma_{t_1,t_0,2} = \gamma_2$ and $B_{t_1,t_0,2} = B_2$.

Return

Error bounds: approximation and simulation errors 2

Simulation errors

$$E_{sf}(d) = (L_1 - 1)\dots(L_d - 1) \sum_{t_1,\dots,t_d \in \{2,3\}} \beta_{t_1,\dots,t_d}$$

$$E_{sapp}(d) = \sum_{s=1}^{d} (L_1 - 1)\cdots(L_s - 1) \sum_{t_1,\dots,t_{s-1} \in \{2,3\}} F_{t_1,\dots,t_{s-1}} \Big(1 - \hat{Q}_{t_1,\dots,t_{s-1},2} + A_{t_1,\dots,t_{s-1},2} + C_{t_1,\dots,t_{s-1},2}\Big)^2$$

where for $2 \leq s \leq d$

$$A_{t_1,\dots,t_{s-1}} = (L_s - 1)\dots(L_d - 1) \sum_{t_s,\dots,t_d \in \{2,3\}} \beta_{t_1,\dots,t_d},\ A_{t_1,\dots,t_d} = \beta_{t_1,\dots,t_d}$$

$$C_{t_1\dots,t_{s-1}} = (L_s - 1)\Big[F_{t_1,\dots,t_{s-1}} \Big(1 - \hat{Q}_{t_1,\dots,t_{s-1},2} + A_{t_1\dots,t_{s-1},2} + C_{t_1\dots,t_{s-1},2}\Big)^2 + \sum_{t_s \in \{2,3\}} C_{t_1\dots,t_s}\Big]$$

Return