

Approximations for the Distribution of Three-dimensional Discrete Scan Statistics

Alexandru Amărioarei
Cristian Preda

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

International Workshop on Applied Probability
11-14 June, 2012, Jerusalem, Israel

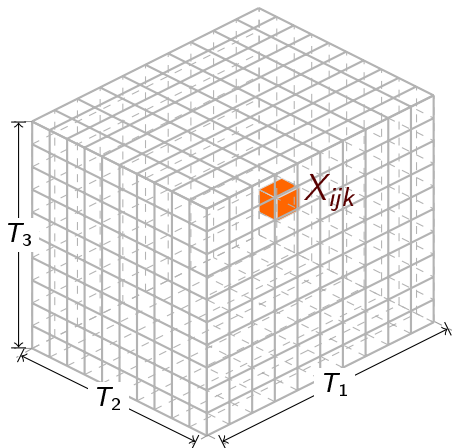
Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

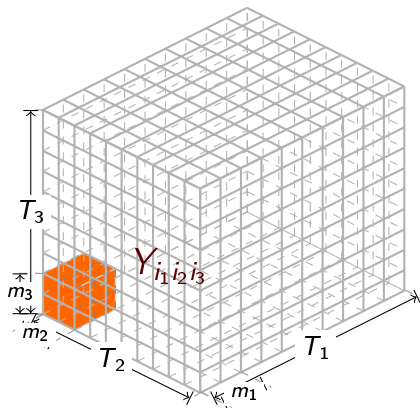
Introducing the Model



Let T_1, T_2, T_3 be positive integers

- Rectangular region
 $\mathcal{R} = [0, T_1] \times [0, T_2] \times [0, T_3]$
- $(X_{ijk})_{\substack{1 \leq i \leq T_1 \\ 1 \leq j \leq T_2 \\ 1 \leq k \leq T_3}}$ i.i.d. integer r.v.'s
 - Bernoulli($\mathcal{B}(1, p)$)
 - Binomial($\mathcal{B}(n, p)$)
 - Poisson($\mathcal{P}(\lambda)$)
- X_{ijk} number of observed events in the elementary subregion
 $r_{ijk} = [i-1, i] \times [j-1, j] \times [k-1, k]$

Defining the Scan Statistic



Let m_1, m_2, m_3 be positive integers

- Define for $1 \leq i_j \leq T_j - m_j + 1$,

$$Y_{i_1 i_2 i_3} = \sum_{i=i_1}^{i_1+m_1-1} \sum_{j=i_2}^{i_2+m_2-1} \sum_{k=i_3}^{i_3+m_3-1} X_{ijk}$$

- The three dimensional scan statistic,

$$S_{m_1, m_2, m_3} = \max_{\substack{1 \leq i_1 \leq T_1 - m_1 + 1 \\ 1 \leq i_2 \leq T_2 - m_2 + 1 \\ 1 \leq i_3 \leq T_3 - m_3 + 1}} Y_{i_1 i_2 i_3}.$$

- Used for testing the null hypotheses of randomness against the alternative hypothesis of clustering

Outline

- 1 Introduction
 - Framework and Model
 - **Problem and Previous Work**
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

Introducing the Problem

Problem

Approximate the distribution of three dimensional discrete scan statistic

$$\mathbb{P}(S_{m_1, m_2, m_3} \leq n).$$

- No exact formulas
- A Poisson approximation for the special case ($n = m_1 m_2 m_3$, Bernoulli model): Darling and Waterman (1986)
- For Bernoulli case: Glaz et al. (2010)
 - Product type approximation
 - Three Poisson approximations

An Animated Illustration of the Scan

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

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 (2000,2007)
 - discrete and continuous two dimensional scan statistic: Haiman and Preda (2002,2006)

Writing the Scan as an Extreme of 1-Dependent R.V.'s

Let $L_j = \frac{T_j}{m_j}$, $j \in \{1, 2, 3\}$ positive integers

- Define for $k \in \{1, 2, \dots, L_3 - 1\}$

$$Z_k = \max_{\substack{1 \leq i_1 \leq (L_1 - 1)m_1 + 1 \\ 1 \leq i_2 \leq (L_2 - 1)m_2 + 1 \\ (k-1)m_3 + 1 \leq i_3 \leq km_3 + 1}} Y_{i_1 i_2 i_3}$$

- $(Z_k)_k$ is 1-dependent and stationary
- Observe

$$S_{m_1, m_2, m_3} = \max_{1 \leq k \leq L_3 - 1} Z_k$$

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - **The Main Tools**
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

Two Main Theorems

Let $(Z_k)_{k \geq 1}$ 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 (Haiman 1999)

For any x such that $\mathbb{P}(Z_1 > x) = 1 - q_1 \leq 0.025$ and any integer $m > 3$,

$$\left| q_m - \frac{6(q_1 - q_2)^2 + 4q_3 - 3q_4}{(1 + q_1 - q_2 + q_3 - q_4 + 2q_1^2 + 3q_2^2 - 5q_1q_2)^m} \right| \leq \bar{\Delta}_1(1 - q_1)^3,$$

$$\left| q_m - \frac{2q_1 - q_2}{[1 + q_1 - q_2 + 2(q_1 - q_2)^2]^m} \right| \leq \bar{\Delta}_2(1 - q_1)^2,$$

- $\bar{\Delta}_1 = 561 + 88m[1 + 124m(1 - q_1)^3]$
- $\bar{\Delta}_2 = 9 + 561(1 - q_1) + 3.3m[1 + 4.7m(1 - q_1)^2]$.

Improved Results

Main Theorem

For x such that $\mathbb{P}(Z_1 > x) = 1 - q_1 \leq \alpha < \frac{4}{27}$ and $m > 3$ we have

$$\left| q_m - \frac{6(q_1 - q_2)^2 + 4q_3 - 3q_4}{(1 + q_1 - q_2 + q_3 - q_4 + 2q_1^2 + 3q_2^2 - 5q_1q_2)^m} \right| \leq \Delta_1(1 - q_1)^3,$$

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

- $\Delta_1 = \Delta_1(\alpha, q_1, m) = \Gamma(\alpha) + mK(\alpha) [1 + 3(1 - q_1)^2]$
- $\Delta_2 = mF(\alpha, q_1, m) = m \left[1 + \frac{3}{m} + K(\alpha)(1 - q_1) + \frac{\Gamma(\alpha)(1 - q_1)}{m} \right].$

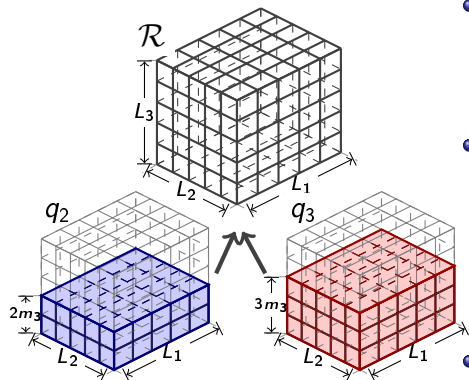
Advantages

- Increased range of applicability
- Sharp bounds values (ex. $\alpha = 0.025$: $561 \rightarrow 162$ and $88 \rightarrow 22$)

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - **The Approximation**
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

First Step Approximation



- Define

$$q_2 = \mathbb{P}(Z_1 \leq n)$$

$$q_3 = \mathbb{P}(Z_1 \leq n, Z_2 \leq n)$$

- If $1 - q_2 \leq \alpha_1 < \frac{4}{27}$ the (first) approximation

$$\mathbb{P}(S \leq n) \approx f(q_2, q_3, L_3 - 1)$$

where $S = S_{m_1, m_2, m_3}$ and

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

- Approximation error

$$(L_3 - 1)F(\alpha_1, L_3 - 1)(1 - q_2)^2$$

with $F(\alpha, \alpha, m) = F(\alpha, m)$.

Approximation for q_2 and q_3

q_2 :

- For $l \in \{1, 2, \dots, L_2 - 1\}$

$$Z_l^{(2)} = \max_{\substack{1 \leq i_1 \leq (L_1 - 1)m_1 + 1 \\ (l-1)m_2 + 1 \leq i_2 \leq lm_2 + 1 \\ 1 \leq i_3 \leq m_3 + 1}} Y_{i_1 i_2 i_3}$$

- $q_2 = \mathbb{P} \left(\max_{1 \leq l \leq L_2 - 1} Z_l^{(2)} \leq n \right)$

- Define

$$q_{22} = \mathbb{P}(Z_1^{(2)} \leq n)$$

$$q_{32} = \mathbb{P}(Z_1^{(2)} \leq n, Z_2^{(2)} \leq n)$$

- Approximation ($1 - q_{22} \leq \alpha_2$)

$$q_2 \approx f(q_{22}, q_{32}, L_2 - 1)$$

-

$$(L_2 - 1)F(\alpha_2, L_2 - 1)(1 - q_{22})^2$$

q_3 :

- For $l \in \{1, 2, \dots, L_2 - 1\}$

$$Z_l^{(3)} = \max_{\substack{1 \leq i_1 \leq (L_1 - 1)m_1 + 1 \\ (l-1)m_2 + 1 \leq i_2 \leq lm_2 + 1 \\ 1 \leq i_3 \leq 2m_3 + 1}} Y_{i_1 i_2 i_3}$$

- $q_3 = \mathbb{P} \left(\max_{1 \leq l \leq L_2 - 1} Z_l^{(3)} \leq n \right)$

- Define

$$q_{23} = \mathbb{P}(Z_1^{(3)} \leq n)$$

$$q_{33} = \mathbb{P}(Z_1^{(3)} \leq n, Z_2^{(3)} \leq n)$$

- Approximation ($1 - q_{23} \leq \alpha_2$)

$$q_3 \approx f(q_{23}, q_{33}, L_2 - 1)$$

-

$$(L_2 - 1)F(\alpha_2, L_2 - 1)(1 - q_{23})^2$$

Illustration of q_{ts} Construction

Last Step (Approximating q_{ts})

Applying again the second part of the Main Theorem...

- For $s, t \in \{2, 3\}$ and $j \in \{1, 2, \dots, L_1 - 1\}$ define

$$Z_j^{(ts)} = \max_{\substack{(j-1)m_1+1 \leq i_1 \leq jm_1+1 \\ 1 \leq i_2 \leq (t-1)m_2+1 \\ 1 \leq i_3 \leq (s-1)m_3+1}} Y_{i_1 i_2 i_3}$$

- Observe

$$q_{ts} = \mathbb{P} \left(\max_{1 \leq j \leq L_1 - 1} Z_j^{(ts)} \leq n \right)$$

- Define for $r, s, t \in \{2, 3\}$,

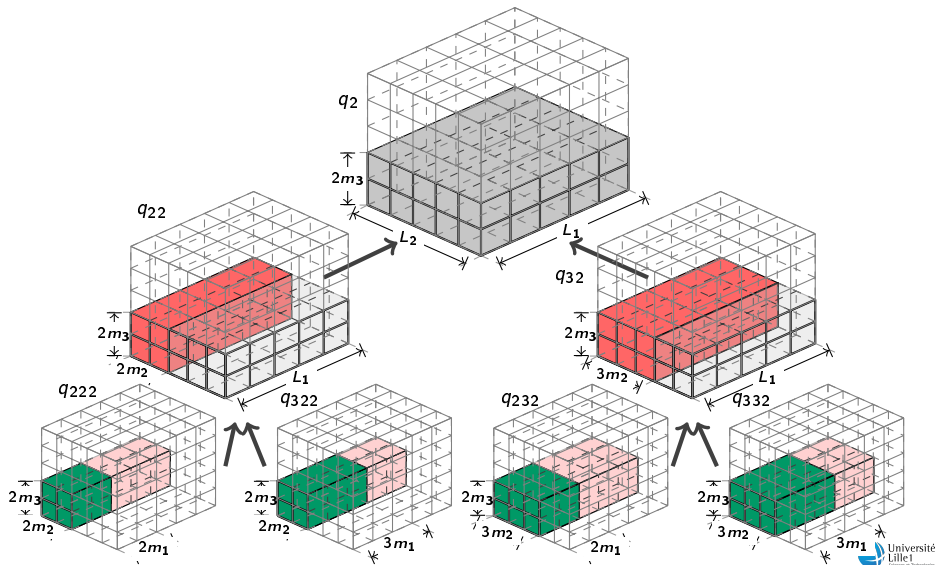
$$q_{rts} = \mathbb{P} \left(\bigcap_{j=1}^{r-1} \{Z_j^{(ts)} \leq n\} \right)$$

- If $1 - q_{2ts} \leq \alpha_3$ then the approximation and the error

$$q_{ts} \approx f(q_{2ts}, q_{3ts}, L_1 - 1)$$

$$(L_1 - 1)F(\alpha_3, L_1 - 1)(1 - q_{2ts})^2$$

An Illustration of the Approximation Chain (q_2)



Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 **Error Bound**
 - **The Approximation Error**
 - The Simulation Error
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

Approximation Error

Define for $t, s \in \{2, 3\}$

$$\begin{cases} \alpha_3 &= 1 - q_3, \alpha_{23} = 1 - q_{23}, \alpha_{233} = 1 - q_{233}, \\ \gamma_{ts} &= f(q_{2ts}, q_{3ts}, L_1 - 1), \gamma_s = f(\gamma_{2s}, \gamma_{3s}, L_2 - 1) \\ F_1 &= F(\alpha_3, L_3 - 1), F_2 = F(\alpha_{23}, L_2 - 1), F_3 = F(\alpha_{233}, L_1 - 1) \end{cases}$$

The approximation error

$$E_{app} = (L_3 - 1)F_1\delta_2^2 + (L_3 - 2)(L_2 - 1)F_2(\delta_{22}^2 + \delta_{23}^2) + \\ + (L_3 - 2)(L_2 - 2)(L_1 - 1)F_3 \left[\sum_{t,s \in \{2,3\}} (1 - q_{2ts})^2 \right]$$

where δ_{22} , δ_{23} , δ_2 are given by

$$\begin{cases} \delta_2 &= 1 - \gamma_2 + (L_2 - 1)F_2\delta_{22} + (L_2 - 2)(L_1 - 1)F_3 [(1 - q_{222})^2 + \\ &+ (1 - q_{232})^2] \\ \delta_{2s} &= 1 - \gamma_{2s} + (L_1 - 1)F_3(1 - q_{22s})^2 \end{cases}$$

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 **Error Bound**
 - The Approximation Error
 - **The Simulation Error**
- 4 Simulation and Numerical Results
 - Simulation
 - Numerical Results
- 5 Further Remarks
- 6 References

Simulation Error for Approximation Formula

If $ITER$ is the number of simulations, we can say, at 95% confidence level,

$$|q_{rts} - \hat{q}_{rts}| \leq 1.96 \sqrt{\frac{\hat{q}_{rts}(1-\hat{q}_{rts})}{ITER}} = \beta_{rts}, \quad r, t, s \in \{2, 3\}$$

where \hat{q}_{rts} is the simulated value.

Define for $t, s \in \{2, 3\}$,

$$\begin{cases} \hat{q}_{ts} &= f(\hat{q}_{2ts}, \hat{q}_{3ts}, L_1 - 1) \\ \hat{q}_s &= f(\hat{q}_{2s}, \hat{q}_{3s}, L_2 - 1) \end{cases}$$

The simulation error corresponding to the approximation formula

$$E_{sf} = (L_1 - 2)(L_2 - 2)(L_3 - 2) \left(\sum_{r,t,s \in \{2,3\}} \beta_{rts} \right)$$

Simulation Error for Approximation Error

In the approximation error formula we consider the transformations

$$\begin{cases} 1 - q_{rts} & \rightarrow 1 - \hat{q}_{rts} + \beta_{rts} = u_{rts} \\ 1 - \gamma_{ts} & \rightarrow 1 - \hat{q}_{ts} + (L_1 - 2)(\beta_{2ts} + \beta_{3ts}) = u_{ts} \\ 1 - \gamma_s & \rightarrow 1 - \hat{q}_s + (L_1 - 2)(L_2 - 2)(\beta_{22s} + \beta_{32s} + \beta_{23s} + \beta_{33s}) = u_s \end{cases}$$

the simulation error corresponding to the approximation error become

$$E_{sapp} = (L_3 - 1)F_1\bar{\delta}_2^2 + (L_3 - 2)(L_2 - 1)F_2(\bar{\delta}_{22}^2 + \bar{\delta}_{23}^2) + (L_3 - 2)(L_2 - 2)(L_1 - 1)F_3(u_{222}^2 + u_{223}^2 + u_{232}^2 + u_{232}^2 + u_{233}^2).$$

where $\bar{\delta}_{22}$, $\bar{\delta}_{23}$, $\bar{\delta}_2$ are given by

$$\begin{cases} \bar{\delta}_{2s} & = u_{2s} + (L_1 - 1)F_3u_{22s}^2 \\ \bar{\delta}_2 & = u_2 + (L_2 - 1)F_2\bar{\delta}_{22} + (L_2 - 2)(L_1 - 1)F_3(u_{222}^2 + u_{232}^2) \end{cases}$$

The total simulation error

$$E_{sim} = E_{sf} + E_{sapp}$$

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results**
 - **Simulation**
 - Numerical Results
- 5 Further Remarks
- 6 References

Simulation of q_{rts}

To simulate q_{rts} we make use of known information using the recurrence

$$q_{rts} = \max\{q_{(r-1)ts}, q_{r(t-1)s}, q_{rt(s-1)}, g_{sum}(r-2, t-2, s-2)\}$$

$$g_{sum}(c_x, c_y, c_z) = \mathbb{P} \left(\begin{array}{l} \max \\ c_x m_1 + 1 \leq i_1 \leq (c_x + 1)m_1 + 1 \\ c_y m_2 + 1 \leq i_2 \leq (c_y + 1)m_2 + 1 \\ c_z m_3 + 1 \leq i_3 \leq (c_z + 1)m_3 + 1 \end{array} Y_{i_1 i_2 i_3} \leq n \right)$$

Outline

- 1 Introduction
 - Framework and Model
 - Problem and Previous Work
- 2 Description of the Method
 - The Key Idea
 - The Main Tools
 - The Approximation
- 3 Error Bound
 - The Approximation Error
 - The Simulation Error
- 4 Simulation and Numerical Results**
 - Simulation
 - Numerical Results**
- 5 Further Remarks
- 6 References

Numerical Results for $\mathbb{P}(S \leq k)$

Comparing with existing results:

Table 1 : $n = 1, p = 0.00005, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Glaz et al. Product type	Our Approximation	Approximation Error	Simulation Error	Total Error
1	0.906980	0.906970	0.909820	0.00918789	0.017163	0.026351
2	0.999540	0.999519	0.999439	0.0000012	0.000483	0.000483

Table 2 : $n = 1, p = 0.0001, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Glaz et al. Product type	Our Approximation	Approximation Error	Simulation Error	Total Error
1	0.685780	0.680843	0.694769	0.23100630	0.250852	0.481859
2	0.996020	0.996203	0.996129	0.00000892	0.001325	0.001334
3	0.999990	0.999980	0.999942	0.00000000	0.000091	0.000091

Numerical Results for $\mathbb{P}(S \leq k)$

Comparing with existing results:

Table 1 : $n = 1, p = 0.00005, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Glaz et al. Product type	Our Approximation	Approximation Error	Simulation Error	Total Error
1	0.906980	0.906970	0.909820	0.00918789	0.017163	0.026351
2	0.999540	0.999519	0.999439	0.00000012	0.000483	0.000483

Table 2 : $n = 1, p = 0.0001, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Glaz et al. Product type	Our Approximation	Approximation Error	Simulation Error	Total Error
1	0.685780	0.680843	0.694769	0.23100630	0.250852	0.481859
2	0.996020	0.996203	0.996129	0.00000892	0.001325	0.001334
3	0.999990	0.999980	0.999942	0.00000000	0.000091	0.000091

Numerical Results for $\mathbb{P}(S \leq k)$

Scanning the same region \mathcal{R} with windows of the same volume but different sizes:

Table 3 :

$n = 1, p = 0.0025, m_1 = 4, m_2 = 4, m_3 = 4, L_1 = 10, L_2 = 10, L_3 = 10, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
4	0.986164	0.98614780	0.00009884	0.00267389	0.00277274
5	0.999568	0.99966216	0.00000008	0.00040217	0.00040226
6	0.999993	0.99998720	0.00000000	0.00002838	0.00002838

Table 4 : $n = 1, p = 0.0025, m_1 = 8, m_2 = 4, m_3 = 2, L_1 = 5, L_2 = 10, L_3 = 20, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
4	0.989160	0.98920297	0.00004532	0.00201212	0.00205744
5	0.999710	0.99953831	0.00000006	0.00035482	0.00035488
6	0.999990	0.99999975	0.00000000	0.00010710	0.00010710

Numerical Results for $\mathbb{P}(S \leq k)$

Scanning the same region \mathcal{R} with windows of the same volume but different sizes:

Table 3 :

$n = 1, p = 0.0025, m_1 = 4, m_2 = 4, m_3 = 4, L_1 = 10, L_2 = 10, L_3 = 10, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
4	0.986164	0.98614780	0.00009884	0.00267389	0.00277274
5	0.999568	0.99966216	0.00000008	0.00040217	0.00040226
6	0.999993	0.99998720	0.00000000	0.00002838	0.00002838

Table 4 : $n = 1, p = 0.0025, m_1 = 8, m_2 = 4, m_3 = 2, L_1 = 5, L_2 = 10, L_3 = 20, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
4	0.989160	0.98920297	0.00004532	0.00201212	0.00205744
5	0.999710	0.99953831	0.00000006	0.00035482	0.00035488
6	0.999990	0.99999975	0.00000000	0.00010710	0.00010710

Numerical Results for $\mathbb{P}(S \leq k)$

Binomial $\mathcal{B}(n, p)$ v.s. Poisson $\mathcal{P}(\lambda)$

Table 5 : $n = 10, p = 0.0025, m_1 = m_2 = m_3 = 4, L_1 = L_2 = L_3 = 20, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.961280	0.96299099	0.00017863	0.00133474	0.00151337
12	0.994650	0.99346695	0.00000220	0.00043587	0.00043807
13	0.999380	0.99976914	0.00000001	0.00012910	0.00012912
14	0.999940	0.99993160	0.00000000	0.00006483	0.00006484

Table 6 : $\lambda = 0.025, m_1 = m_2 = m_3 = 4, L_1 = L_2 = L_3 = 20, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.959000	0.95595046	0.00024518	0.01633211	0.01657730
12	0.994625	0.99452042	0.00000362	0.00553831	0.00554193
13	0.999550	0.99923331	0.00000024	0.00265313	0.00265337
14	0.999975	1	--	--	--

Numerical Results for $\mathbb{P}(S \leq k)$

Binomial $\mathcal{B}(n, p)$ v.s. Poisson $\mathcal{P}(\lambda)$

Table 5 : $n = 10, p = 0.0025, m_1 = m_2 = m_3 = 4, L_1 = L_2 = L_3 = 20, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.961280	0.96299099	0.00017863	0.00133474	0.00151337
12	0.994650	0.99346695	0.00000220	0.00043587	0.00043807
13	0.999380	0.99976914	0.00000001	0.00012910	0.00012912
14	0.999940	0.99993160	0.00000000	0.00006483	0.00006484








Table 6 : $\lambda = 0.025, m_1 = m_2 = m_3 = 4, L_1 = L_2 = L_3 = 20, ITER = 10^7$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.959000	0.95595046	0.00024518	0.01633211	0.01657730
12	0.994625	0.99452042	0.00000362	0.00553831	0.00554193
13	0.999550	0.99923331	0.00000024	0.00265313	0.00265337
14	0.999975	1	--	--	--

Remarks

We need to reduce the simulation error:

- increasing the iterations number
- develop faster algorithms
- use variance reduction techniques

-  Glaz, J., Naus, J., Wallenstein, S.: Scan statistic. *Springer* (2001).
-  Darling, R., Waterman, M.: Approximations for three dimensional scan statistic. *SIAM J. Appl Math* **46** (1986), 118–132.
-  Glaz, J., Guerriero, M., Sen, R.: Approximations for three dimensional scan statistic. *Methodol Comput Appl Probab* **12** (2010), 731–747.
-  Haiman, G.: First passage time for some stationary sequence. *Stoch Proc Appl* **80** (1999), 231–248.
-  Haiman, G.: Estimating the distribution of scan statistics with high precision. *Extremes* **3** (2000), 349–361.
-  Haiman, G., Preda, C.: A new method for estimating the distribution of scan statistics for a two-dimensional Poisson process. *Methodol Comput Appl Probab* **4** (2002), 393–407.
-  Haiman, G., Preda, C.: Estimation for the distribution of two-dimensional scan statistics. *Methodol Comput Appl Probab* **8** (2006), 373–381.

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

α	$K(\alpha)$	$\Gamma(\alpha)$
0.1	79.6678	1471.62
0.075	43.229	454.412
0.05	29.0284	245.873
0.025	21.5672	161.737
0.01	18.5949	132.618
0.005	17.765	124.924
0.0025	17.3749	121.373

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


How to compute $K(\alpha)$ and $\Gamma(\alpha)$?

For $\epsilon = 0.0001$ and $\alpha < \frac{4}{27} - \epsilon$, let's denote with $t_2(\alpha)$ the second largest solution of the equation $\alpha t^3 - t + 1 = 0$ and define $l = t_2(\alpha)^3 + \epsilon$:

$$K(\alpha) = \frac{\frac{11+2\alpha-5\alpha^2}{(1-\alpha)^2} + 2l(1+4\alpha)(1+l\alpha) \left\{ \frac{4}{[1-\alpha(1+l\alpha)^2]^3} - 1 \right\}}{1-\alpha(1+l\alpha)^2 \left\{ \frac{2}{[1-\alpha(1+l\alpha)^2]^2} + \frac{1}{1-\alpha(1+l\alpha)^2} \right\}}$$

For the formula expressing $\Gamma(\alpha)$, define first

$$L(\alpha) = 6(7 + 3\alpha) + (1 + 2\alpha + \alpha^2)P(\alpha)$$

$$P(\alpha) = 19 + 36\alpha + 27\alpha^2 + 27\alpha^3 + 3(1 + \alpha + 3\alpha^2)^2 K(\alpha) + 3\alpha^3(1 + \alpha + 3\alpha^2)K^2(\alpha) + \alpha^6 K^3(\alpha).$$

and if we denote by $\eta = 1 + l\alpha$ and

$$E(\alpha) = \frac{\eta^5 [1 + \alpha(\eta - 1)][1 + \eta + \eta^2 + \alpha\eta^2(\eta - 3)][1 + \eta + \alpha\eta(\eta - 2)]^4}{(1 - \alpha\eta^2)^4 \{ (1 - \alpha\eta^2)^2 - \alpha\eta^2 [1 + \eta + \alpha\eta(\eta - 2)]^2 \}}.$$

$$\Gamma(\alpha) = L(\alpha) + E(\alpha)$$