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

SPSR Conference
April 27, 2012

Outline

1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - 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

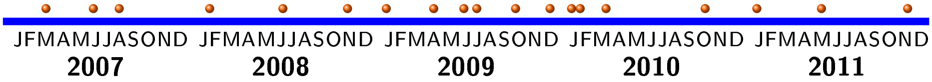
1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - The Approximation
3. Error Bound
   - The Approximation Error
   - The Simulation Error
4. Simulation and Numerical Results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

An intuitive example

Observation of disease cases over a period of time

- Number of years 5
- Number of observed disease cases $N = 19$

Can you observe something unusual?

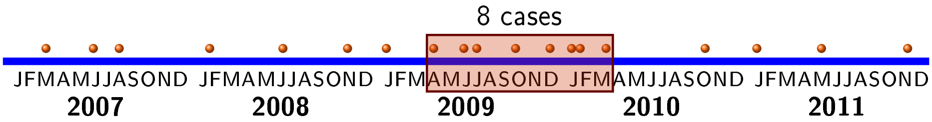


An intuitive example

Observation of disease cases over a period of time

- Number of years 5
- Number of observed disease cases $N = 19$

Can you observe something unusual?

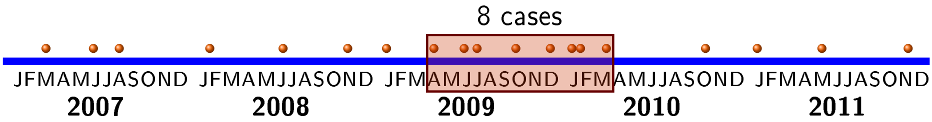


An intuitive example

Observation of disease cases over a period of time

- Number of years 5
- Number of observed disease cases $N = 19$

Can you observe something unusual?



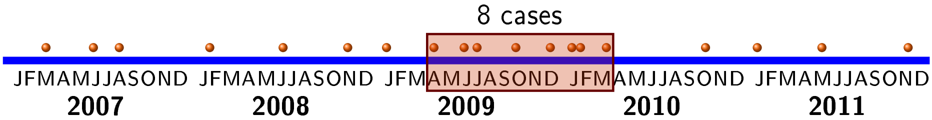
- The epidemiologist observe a 1 **year period** (April 09- April 10) with 8 **cases**: 42%!

An intuitive example

Observation of disease cases over a period of time

- Number of years 5
- Number of observed disease cases $N = 19$

Can you observe something unusual?



- The epidemiologist observe a 1 **year period** (April 09- April 10) with 8 **cases**: 42%!

How unusual is to have during a 1 year period as much as 8 cases ?

The answer

(1) First approach:

- X = number of cases falling in (April 09-April 10)
- $X \sim Bin(19, 0.2)$

$$\mathbb{P}(X \geq 8) = 0.023$$

The answer

(1) First approach:

- X = number of cases falling in (April 09-April 10)
- $X \sim Bin(19, 0.2)$

$$\mathbb{P}(X \geq 8) = 0.023$$

Conclusion: atypical situation !!!

The answer

(1) First approach:

- X = number of cases falling in (April 09-April 10)
- $X \sim Bin(19, 0.2)$

$$\mathbb{P}(X \geq 8) = 0.023$$

Conclusion: atypical situation !!!

However, this is **not** the answer to the question: the 1 year period is **not fixed** but identified after the scanning process!

The answer

(1) First approach:

- X = number of cases falling in (April 09-April 10)
- $X \sim Bin(19, 0.2)$

$$\mathbb{P}(X \geq 8) = 0.023$$

Conclusion: atypical situation !!!

However, this is **not** the answer to the question: the 1 year period is **not fixed** but identified after the scanning process!

(2) The scan statistic approach:

S = max number of cases over **any continuous** 1 year period in $[0, T]$

Thus the answer to the question is

$$\mathbb{P}(S \geq 8) = 0.379$$

Conclusion: **not unusual situation!!!**

Outline

1 Introduction
- What is the scan statistic?
- Framework and Model
- Problem and Previous Work

2 Description of the Method
- The Main Tools
- The Key Idea
- The Approximation

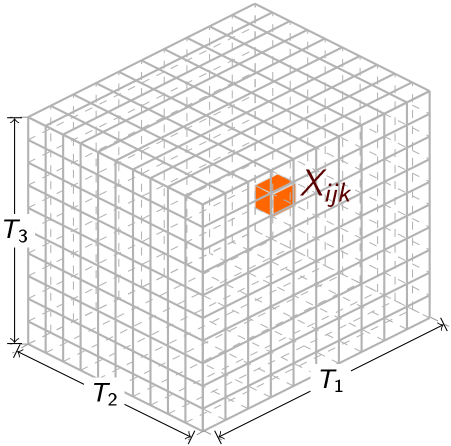
3 Error Bound
- The Approximation Error
- The Simulation Error

4 Simulation and Numerical Results
- Simulation
- Numerical Results

5 Further Remarks

6 References

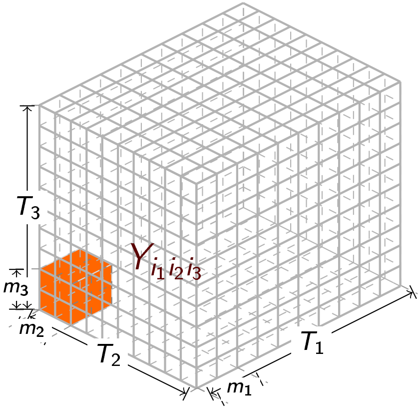
Introducing the Model



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
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - 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

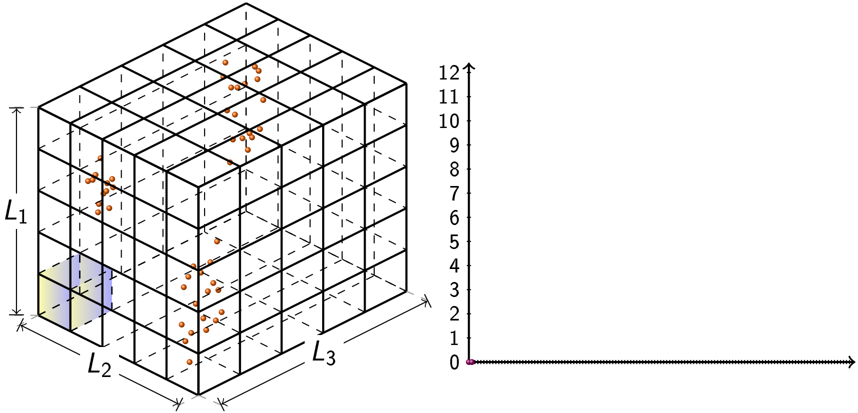
Problem

Approximate the distribution of three dimensional discrete scan statistic

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

- 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



Number of points: 0

Outline

1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - 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 $q_m = q_m(x) = \mathbb{P}(\max(Z_1, \ldots, Z_m) \leq x)$, 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 $m > 3$,

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

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

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

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

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

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

- $\Delta_1 = \Delta_1(\alpha, q_1, m) = \Gamma(\alpha) + mK(\alpha)\left[1 + 3(1 - q_1)^2\right]$
- $\Delta_2 = \Delta_2(\alpha, q_1, m) = 9 + \Gamma(\alpha)(1 - q_1) + m\left[1 + K(\alpha)(1 - q_1)\right]$.

Advantages

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

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

Outline

1 Introduction
- What is the scan statistic?
- Framework and Model
- Problem and Previous Work

2 Description of the Method
- The Main Tools
- The Key Idea
- The Approximation

3 Error Bound
- The Approximation Error
- The Simulation Error

4 Simulation and Numerical Results
- Simulation
- Numerical Results

5 Further Remarks

6 References

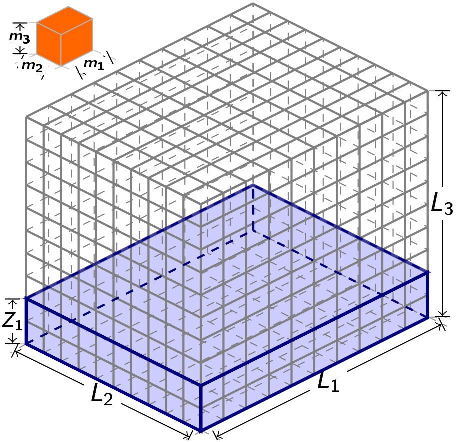
A Different Approach

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 $k \in \{1, 2, \ldots, L_3 - 1\}$

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

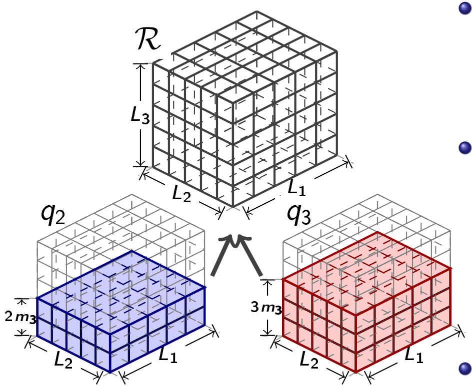
- $(Z_k)_k$ is 1-dependent and stationary
- Observe

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

Outline

1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - 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



We adopt a mixed version...

- 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
$$f(x, y, m) = \frac{x+|x-y|}{[1+|x-y|+2(x-y)^2]^m}$$
- Approximation error
$$\Delta_2(\alpha_1, q_2, L_3 - 1)(1 - q_2)^2$$

Approximation for q_2 and q_3

q_2 :

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

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

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

- Define

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

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

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

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

- $$\Delta_2(\alpha_2, q_{22}, L_2 - 1)(1 - q_{22})^2$$

q_3 :

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

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

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

- Define

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

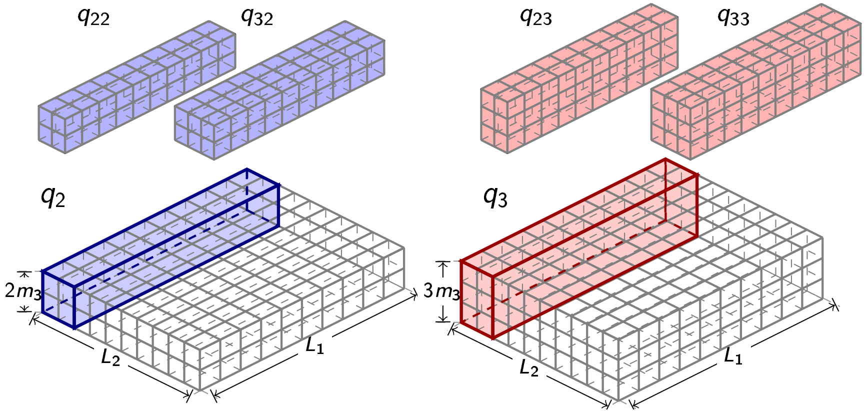
$$q_{33} = \mathbb{P}(Z_1^{(3)} \le n, Z_2^{(3)} \le n)$$

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

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

- $$\Delta_2(\alpha_2, q_{23}, L_2 - 1)(1 - q_{23})^2$$

Illustration of q_{ts} Construction



Last Step (Approximating q_{ts})

Applying the first 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 $r \in \{2, 3, 4, 5\}$ and $s, t \in \{2, 3\}$,

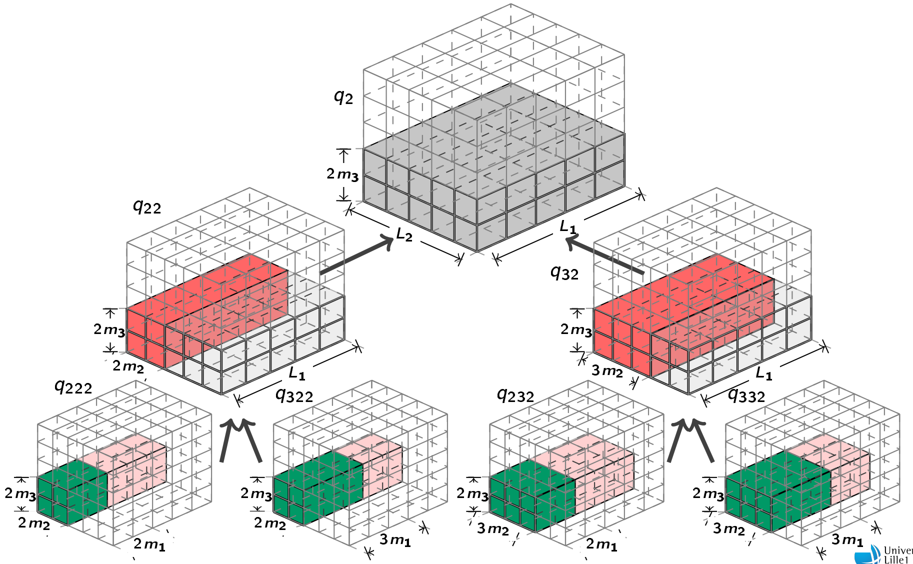
$$q_{rts} = \mathbb{P}\left(\cap_{j=1}^{r-1} \{ Z_j^{(ts)} \leq n \} \right)$$

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

$$q_{ts} \approx h\left(q_{2ts}, q_{3ts}, q_{4ts}, q_{5ts}, L_1 - 1\right) \qquad \Delta_1(\alpha_3, q_{2ts}, L_1 - 1)(1 - q_{2ts})^3$$

where $h(x, y, z, t, m) = \frac{6(x-y)^2+z+3|z-t|}{[1+2(x-y)^2+|x-y|(1-y)+|z-t|]^m}$

An Illustration of the Approximation Chain (q_2)



Outline

1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - 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 $t, s \in \{2, 3\}$

$$\begin{cases} \alpha_3 & = & 1 - q_3, \ \alpha_{23} = 1 - q_{23}, \ \alpha_{233} = 1 - q_{233}, \\ \gamma_{ts} & = & h(q_{2ts}, q_{3ts}, q_{4ts}, q_{5ts}, L_1 - 1) \\ \gamma_s & = & f(\gamma_{2s}, \gamma_{3s}, L_2 - 1) \end{cases}$$

The approximation error

$$\begin{aligned} E_{app} &= (L_3 - 1)a_1(1 - \gamma_2)^2 + (L_3 - 1)(L_2 - 1)\left[a_2(1 - \gamma_{22})^2 + \right. \\ &\quad + a_3(1 - \gamma_{23})^2\Big] + (L_3 - 1)(L_2 - 1)(L_1 - 1)\left[a_4(1 - q_{222})^3 + \right. \\ &\quad \left. + a_5(1 - q_{223})^3 + a_6(1 - q_{232})^3 + a_7(1 - q_{233})^3\right]. \end{aligned}$$

where a_j, $j \in \{1, 2, \ldots, 7\}$, are functions of $\alpha_3, \alpha_{23}, \alpha_{233}$ and L_1, L_2, L_3.

▸ Computation of the coefficients

Université Lille1 Sciences et Technologies

Outline

1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - 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,

$$|q_{rts} - \hat{q}_{rts}| \leq 1.96\sqrt{\frac{\hat{q}_{rts}(1-\hat{q}_{rts})}{ITER}},\ r \in \{2,3,4,5\},\ t,s \in \{2,3\}$$

where $\hat{q}_{rts}$ is the simulated value.
Define for $t,s \in \{2,3\}$,

$$\begin{cases} \hat{q}_{ts} &= h(\hat{q}_{2ts}, \hat{q}_{3ts}, \hat{q}_{4ts}, \hat{q}_{5ts}, L_1 - 1) \\ \hat{q}_{s} &= f(\hat{q}_{2s}, \hat{q}_{3s}, L_2 - 1) \\ B_{ts} &= S_1(\hat{q}_{2ts}, \hat{q}_{3ts}, \hat{q}_{4ts}, \hat{q}_{5ts}, L_1 - 1) \\ B_{s} &= S_2(\hat{q}_{2s}, \hat{q}_{3s}, B_{2s}, B_{3s}, L_2 - 1) \end{cases}$$

The simulation error corresponding to the approximation formula

$$E_{sf} = (L_1 - 1)(\hat{q}_2 + |\hat{q}_2 - \hat{q}_3|)(B_2 + B_3)(1 + 2B_2 + 2B_3 + 4|\hat{q}_2 - \hat{q}_3|)$$

where the functions S_1 and S_2 are defined in the appendix.

Details for S_1 and S_2

Simulation Error for Approximation Error

In the approximation error formula we consider the transformations

$$\begin{cases} 1 - q_{rts} & \rightarrow & 1 - \hat{q}_{rts} + g(\hat{q}_{rts}) = u_{rts} \\ 1 - \gamma_{ts} & \rightarrow & 1 - \hat{q}_{ts} + B_{ts} = u_{ts} \\ 1 - \gamma_{s} & \rightarrow & 1 - \hat{q}_{s} + B_{s} = u_{s} \end{cases}$$

the simulation error corresponding to the approximation error become

$$\begin{aligned} E_{sapp} &= (L_3 - 1)\bar{a}_1 u_2^2 + (L_3 - 1)(L_2 - 1)\left(\bar{a}_2 u_{22}^2 + \bar{a}_3 u_{23}^2\right) \\ &+ (L_3 - 1)(L_2 - 1)(L_1 - 1)\left(\bar{a}_4 u_{222}^3 + \bar{a}_5 u_{223}^3 + \bar{a}_6 u_{232}^3 + \bar{a}_7 u_{233}^3\right) \end{aligned}$$

The total simulation error

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

▸ Computation of the coefficients

Université Lille1 Sciences et Technologies

Outline

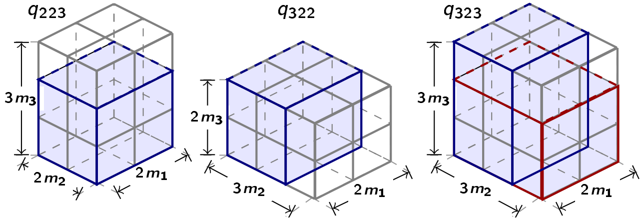
1. Introduction
   - What is the scan statistic?
   - Framework and Model
   - Problem and Previous Work
2. Description of the Method
   - The Main Tools
   - The Key Idea
   - The Approximation
3. Error Bound
   - The Approximation Error
   - The Simulation Error
4. Simulation and Numerical Results
   - 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

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

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



Outline

1 Introduction
- What is the scan statistic?
- Framework and Model
- Problem and Previous Work

2 Description of the Method
- The Main Tools
- The Key Idea
- 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: $n = 1, p = 0.00005, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 4 \times 10^6$

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	0.905566	0.00863037	0.539903	0.548534
2	0.999540	0.999519	0.999469	0.00000008	0.028854	0.028854

Table 2: $n = 1, p = 0.0001, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 4 \times 10^6$

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	0.697293	0.24890871	1.684443	1.933352
2	0.996020	0.996203	0.996125	0.00000788	0.086019	0.086027
3	0.999990	0.999980	0.999888	0.00000001	0.001469	0.001469

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

Comparing with existing results:

Table 1: $n = 1, p = 0.00005, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 4 \times 10^6$

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	0.905566	0.00863037	0.539903	0.548534
2	0.999540	0.999519	0.999469	0.00000008	0.028854	0.028854

Table 2: $n = 1, p = 0.0001, m_1 = m_2 = m_3 = 5, L_1 = L_2 = L_3 = 10, ITER = 4 \times 10^6$

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	0.697293	0.24890871	1.684443	1.933352
2	0.996020	0.996203	0.996125	0.00000788	0.086019	0.086027
3	0.999990	0.999980	0.999888	0.00000001	0.001469	0.001469

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	0.985580	0.98569657	0.00009120	0.10078854	0.10087973
5	0.999410	0.99957819	0.00000010	0.01878755	0.01878765
6	0.999990	0.99999510	0.00000001	0.00236660	0.00236660

Table 4:
$n = 1, p = 0.0025, m_1 = 8, m_2 = 4, m_3 = 2, L_1 = 5, L_2 = 10, L_3 = 20, ITER = 4 \times 10^6$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
4	0.989160	0.98813583	0.00003107	0.17561572	0.17564680
5	0.999710	0.99975303	0.00000002	0.02850348	0.02850350
6	0.999990	0.99999975	0.00000000	0.00633269	0.00633269

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	0.985580	0.98569657	0.00009120	0.10078854	0.10087973
5	0.999410	0.99957819	0.00000010	0.01878755	0.01878765
6	0.999990	0.99999510	0.00000001	0.00236660	0.00236660

Table 4:
$n = 1, p = 0.0025, m_1 = 8, m_2 = 4, m_3 = 2, L_1 = 5, L_2 = 10, L_3 = 20, ITER = 4 \times 10^6$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
4	0.989160	0.98813583	0.00003107	0.17561572	0.17564680
5	0.999710	0.99975303	0.00000002	0.02850348	0.02850350
6	0.999990	0.99999975	0.00000000	0.00633269	0.00633269

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

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

Table 5: $n = 10, p = 0.0025, m_1 = m_2 = m_3 = 4, L_1 = L_2 = L_3 = 20, ITER = 4 \times 10^6$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.961280	0.96875441	0.00022516	0.39557144	0.39579660
12	0.994650	0.99582741	0.00000437	0.14665738	0.14666175
13	0.999380	0.99976914	0.00000014	0.05806762	0.05806776
14	0.999940	0.99999820	0.00000001	0.00723765	0.00723765

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

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.959000	0.95570144	0.00026601	0.66780767	0.66807369
12	0.994625	0.99359316	0.00000362	0.23978257	0.23978618
13	0.999550	0.99854320	0.00000008	0.06594544	0.06594552
14	0.999975	0.99992800	0.00000001	0.00884761	0.00884762

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

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

Table 5: $n = 10, p = 0.0025, m_1 = m_2 = m_3 = 4, L_1 = L_2 = L_3 = 20, ITER = 4 \times 10^6$

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.961280	0.96875441	0.00022516	0.39557144	0.39579660
12	0.994650	0.99582741	0.00000437	0.14665738	0.14666175
13	0.999380	0.99976914	0.00000014	0.05806762	0.05806776
14	0.999940	0.99999820	0.00000001	0.00723765	0.00723765

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

k	$\hat{\mathbb{P}}(S \leq k)$	Our Approximation	Approximation Error	Simulation Error	Total Error
11	0.959000	0.95570144	0.00026601	0.66780767	0.66807369
12	0.994625	0.99359316	0.00000362	0.23978257	0.23978618
13	0.999550	0.99854320	0.00000008	0.06594544	0.06594552
14	0.999975	0.99992800	0.00000001	0.00884761	0.00884762

Final Observations

- Increasing the iterations number we can obtain better error bounds
- The need of faster algorithms for simulation
- The method works in the same fashion for any law of X_{ijk}
- We prepare the continuous three dimensional scan statistic for a Poisson process

- 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 $\Gamma(\alpha)$

Return

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

For $\epsilon = 0.0001$ and $\alpha < \frac{4}{27} - \epsilon$, lets 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

$$\begin{aligned} L(\alpha) &= 6(7+3\alpha) + (1+2\alpha+\alpha^2)P(\alpha) \\ P(\alpha) &= 19 + 36\alpha + 27\alpha^2 + 27\alpha^3 + 3(1+\alpha+3\alpha^2)^2 K(\alpha) \\ &\quad + 3\alpha^3(1+\alpha+3\alpha^2)K^2(\alpha) + \alpha^6 K^3(\alpha). \end{aligned}$$

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

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

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

The Computation of the Approximation Error Coefficients

If $\alpha_3 = 1 - q_3,\ \alpha_{23} = 1 - q_{23},\ \alpha_{233} = 1 - q_{233},$

$$\begin{cases} \Delta_1(\alpha,\alpha,m) & = & \Gamma(\alpha) + mK(\alpha)\left[1 + 3\alpha^2\right] \\ \Delta_2(\alpha,\alpha,m) & = & 9 + \Gamma(\alpha)\alpha + m\left[1 + K(\alpha)\alpha\right] \end{cases}$$

Denote by

$$\begin{cases} G_1 & = & \frac{\Delta_2(\alpha_3,\alpha_3,L_3-1)}{L_3-1}, \\ G_2 & = & \frac{\Delta_2(\alpha_{23},\alpha_{23},L_2-1)}{L_2-1}, \\ G_3 & = & \frac{\Delta_1(\alpha_{233},\alpha_{233},L_1-1)}{L_1-1}, \\ y_1 & = & \alpha_3 + \Delta_2(\alpha_{23},\alpha_{23},L_2-1)\alpha_{23}^2, \\ y_2 & = & \alpha_{23} + \Delta_1(\alpha_{233},\alpha_{233},L_1-1)\alpha_{233}^3, \\ e_1 & = & G_3\left[\frac{2}{L_2-1} + (1+\alpha_{23})(1+2\alpha_{23}+2y_2\right], \\ e_2 & = & G_3\left[\frac{1}{L_2-1} + (1+\alpha_{23})(1+2\alpha_{23}+2y_2)\right] \end{cases}$$

◂ Return

The Computation of the Approximation Error Coefficients

$$
\begin{aligned}
a_1 &= G_1, \\
a_2 &= G_2\left[(\alpha_3+y_1)G_1+\frac{2}{L_3-1}+(1+\alpha_3)(1+2\alpha_3+2y_1)\right], \\
a_3 &= G_2\left[\frac{1}{L_3-1}+(1+\alpha_3)(1+2\alpha_3+2y_1)\right], \\
a_4 &= e_1\left[a_1(1-\gamma_2+y_1)+\frac{2}{L_3-1}+(1+y_1)(1+2y_1+2|\gamma_2-\gamma_3|)\right] \\
&\quad +G_3 a_2(\alpha_{23}+y_2), \\
a_5 &= G_3(\alpha_{23}+y_2)a_3+e_1\left[\frac{1}{L_3-1}+(1+y_1)(1+2y_1+2|\gamma_2-\gamma_3|)\right], \\
a_6 &= e_2\left[a_1(1-\gamma_2+y_1)+\frac{2}{L_3-1}+(1+y_1)(1+2y_1+2|\gamma_2-\gamma_3|)\right], \\
a_7 &= e_2\left[\frac{1}{L_3-1}+(1+y_1)(1+2y_1+2|\gamma_2-\gamma_3|)\right].
\end{aligned}
$$

Defining the Functions S_1 and S_2

If we denote with

$$\begin{cases} g(x) & = & 1.96\sqrt{\frac{x(1-x)}{ITER}} \\ \nu(x,y,z,t) & = & 6(x-y)^2 + z + 3|z-t| \end{cases}$$

We define the functions S_1 and S_2 as follows:

$$\begin{aligned} S_1(x,y,z,t,m) &= 2(3+m\nu(x,y,z,t))(g(x)+g(y))\left[g(x)+g(y)+\right. \\ &\left. +\,2|x-y|\right] + m\nu(x,y,z,t)(1-y+g(y))+ \\ &+4g(x)+g(y)(3+|x-y|)+g(z)+g(t) \\ S_2(x,y,u,v,m) &= m(x+|x-y|)(u+v)(1+2u+2v+4|x-y|)+ \\ &+2u+v \end{aligned}$$

Return

Coefficients of the Simulation Error (E_{sapp})

$$
\begin{aligned}
\bar{a}_1 &= G_1,\\
\bar{a}_2 &= G_2\left[(\alpha_3+y_1)G_1+\frac{2}{L_3-1}+(1+\alpha_3)(1+2\alpha_3+2y_1)\right],\\
\bar{a}_3 &= G_2\left[\frac{1}{L_3-1}+(1+\alpha_3)(1+2\alpha_3+2y_1)\right],\\
\bar{a}_4 &= e_1\left[a_1(u_2+y_1)+\frac{2}{L_3-1}+(1+y_1)(1+2y_1+2|u_2-u_3|)\right]\\
&\quad +G_3a_2(\alpha_{23}+y_2),\\
\bar{a}_5 &= G_3(\alpha_{23}+y_2)a_3+e_1\left[\frac{1}{L_3-1}+(1+y_1)(1+2y_1+2|u_2-u_3|)\right],\\
\bar{a}_6 &= e_2\left[a_1(u_2+y_1)+\frac{2}{L_3-1}+(1+y_1)(1+2y_1+2|u_2-u_3|)\right],\\
\bar{a}_7 &= e_2\left[\frac{1}{L_3-1}+(1+y_1)(1+2y_1+2|u_2-u_3|)\right].
\end{aligned}
$$

Return