Approximations for the distribution of the discrete scan statistics when the scanning window has arbitrary shape

Alexandru Amărioarei

National Institute of Research and Development for Biological Sciences

The 19^{th} Conference of the Romanian Society of Probability and Statistics
27-28 May, 2016, București, România

Outline

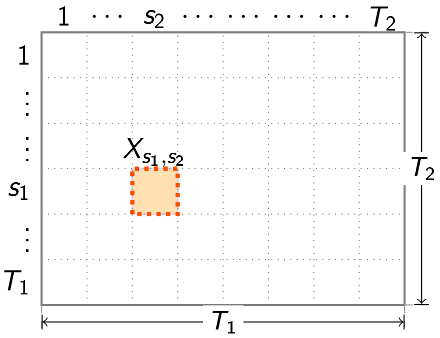
1. Introduction
   - Framework
   - Problem
2. Methodology
   - Approximation
   - Simulation methods: Normal data
3. Simulation study
   - Numerical examples
   - Power
4. References

Outline

Definitions and notations

Preliminary notations

Let T_1, T_2 be positive integers



- Rectangular region $\mathcal{R}_2 = [0, T_1] \times [0, T_2]$
- $(X_{s_1,s_2})_{\substack{1 \leq s_1 \leq T_1 \\ 1 \leq s_2 \leq T_2}}$ i.i.d. r.v.'s
  - Bernoulli($\mathcal{B}(1, p)$)
  - Binomial($\mathcal{B}(n, p)$)
  - Poisson($\mathcal{P}(\lambda)$)
  - Normal($\mathcal{N}(\mu, \sigma^2)$)
- X_{s_1,s_2} number of observed events in the elementary subregion $r_{s_1,s_2} = [s_1 - 1, s_1] \times [s_2 - 1, s_2]$

Two dimensional scan statistic

Let $2 \leq m_s \leq T_s$, $s \in \{1, 2\}$ be positive integers

- Define for $1 \leq i_s \leq T_s - m_s + 1$ and $1 \leq j_s \leq m_s$ the 2-way tensor $\mathcal{X}_{i_1,i_2} \in \mathbb{R}^{m_1 \times m_2}$,

$$\mathcal{X}_{i_1,i_2}(j_1, j_2) = X_{i_1+j_1-1, i_2+j_2-1}$$

- Take $\mathcal{S} : \mathbb{R}^{m_1 \times m_2} \to \mathbb{R}$ to be a measurable real valued function (*score function*) and define

$$Y_{i_1,i_2}(\mathcal{S}) = \mathcal{S}\left(\mathcal{X}_{i_1,i_2}\right)$$

Definition

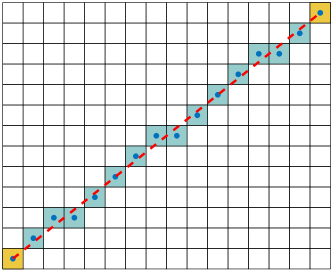
The two dimensional scan statistic with score function $\mathcal{S}$ is defined by

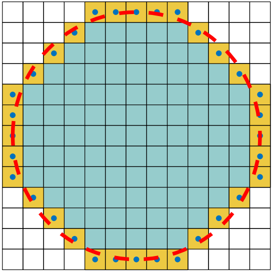
$$S_{m_1,m_2}(T_1, T_2; \mathcal{S}) = \max_{\substack{1 \leq i_1 \leq T_1 - m_1 + 1 \\ 1 \leq i_2 \leq T_2 - m_2 + 1}} Y_{i_1,i_2}(\mathcal{S})$$

Shape of the scanning window

Let G be the geometrical shape of the scanning window (rectangle, quadrilateral, ellipse, etc.) and $\tilde{G}$ be its corresponding discrete form.

- Rasterization algorithms (omputer vision): continuous shape $\rightarrow$ discrete shape
  - Line - Bresenham line algorithm ([Bresenham, 1965])
  - Circle - Bresenham circle algorithm ([Bresenham, 1977])
  - Bezier curves - [Foley, 1995]

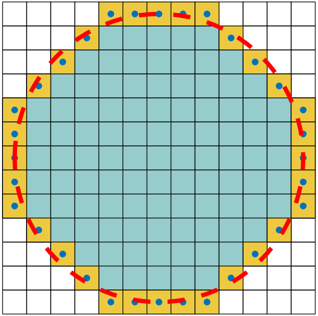




Shape of the scanning window

To each discrete shape $\tilde{G}$ it corresponds an unique matrix (2-way tensor) $A(G) = A\left(\tilde{G}\right)$ (of smallest size) with 0 and 1 entries (1 if there is a point and 0 otherwise):

$\tilde{G}$



$A\left(\tilde{G}\right)$

				1	1	1	1	1				
			1	1	1	1	1	1	1			
		1	1	1	1	1	1	1	1	1		
	1	1	1	1	1	1	1	1	1	1	1	
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	
		1	1	1	1	1	1	1	1	1		
			1	1	1	1	1	1	1			
				1	1	1	1	1				

Arbitrary window scan statistic

Let G be a geometric shape and $A = A(G)$ its corresponding $\{0,1\}$ matrix of size $m_1 \times m_2$.

- Define the score function $\mathcal{S}$ associated to the shape G by

$$\mathcal{S}\left(\mathcal{X}_{i_1,i_2}\right) = A \circ \mathcal{X}_{i_1,i_2} = \sum_{s_1=i_1}^{i_1+m_1-1} \sum_{s_2=i_2}^{i_2+m_2-1} A(s_1 - i_1 + 1, s_2 - i_2 + 1) X_{s_1,s_2}$$

Remark

If, in particular, the shape G is a rectangle of size $m_1 \times m_2$ than its corresponding $\{0,1\}$ matrix of the same size has all the entries equal to 1 so the score function

$$\mathcal{S}\left(\mathcal{X}_{i_1,i_2}\right) = \sum_{s_1=i_1}^{i_1+m_1-1} \sum_{s_2=i_2}^{i_2+m_2-1} X_{s_1,s_2}$$

is the *classical* rectangular window of the two dimensional scan statistics.

Outline

Problem and related work

Objective

Find a good estimate for the distribution of the two dimensional discrete scan statistic with score function $\mathcal{S}$

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

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

Previous work:

- Continuous scan statistics
  - Rectangles: [Loader, 1991], [Glaz et al., 2001], [Glaz et al., 2009]
  - Circles: [Anderson and Titterington, 1997]
  - Triangles, ellipses and other convex shapes: [Alm, 1983, Alm, 1997, Alm, 1998], [Tango and Takahashi, 2005], [Assunção et al., 2006]
- Discrete scan statistics
  - **No results !**

Outline

Approximation methodology for the general scan statistic

Approximation and error bounds

Theorem (Generalization of [Amărioarei, 2014])

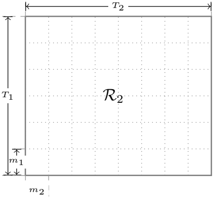
Let $t_1, t_2 \in \{2,3\}$, $Q_{t_1,t_2} = \mathbb{P}\left(S_{\mathbf{m}}\left(t_1(m_1-1), t_2(m_2-1); \mathcal{S}\right) \leq \tau\right)$ and $L_s = \left\lfloor \frac{T_s}{m_s - 1} \right\rfloor$, $s \in \{1,2\}$. If $\hat{Q}_{t_1,t_2}$ is an estimate of Q_{t_1,t_2}, $\left|\hat{Q}_{t_1,t_2} - Q_{t_1,t_2}\right| \leq \beta_{t_1,t_2}$ and τ is such that $1 - \hat{Q}_{2,2}(\tau) \leq 0.1$ then

$$\left|\mathbb{P}\left(S_{\mathbf{m}}(\mathbf{T};\mathcal{S}) \leq \tau\right) - \left(2\hat{Q}_2 - \hat{Q}_3\right)\left[1 + \hat{Q}_2 - \hat{Q}_3 + 2(\hat{Q}_2 - \hat{Q}_3)^2\right]^{1-L_1}\right| \leq E_{sf} + E_{sapp},$$

where, for $t \in \{2,3\}$

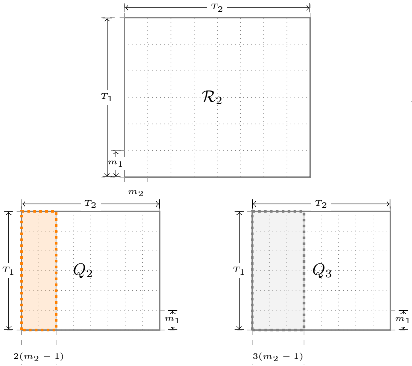
$$\begin{aligned}
\hat{Q}_t &= \left(2\hat{Q}_{t,2} - \hat{Q}_{t,3}\right)\left[1 + \hat{Q}_{t,2} - \hat{Q}_{t,3} + 2(\hat{Q}_{t,2} - \hat{Q}_{t,3})^2\right]^{1-L_2} \\
E_{sf} &= (L_1 - 1)(L_2 - 1)\left(\beta_{2,2} + \beta_{2,3} + \beta_{3,2} + \beta_{3,3}\right) \\
E_{sapp} &= (L_1 - 1)\left[F_1\left(1 - \hat{Q}_2 + A_2 + C_2\right)^2 + (L_2 - 1)(F_2 C_2 + F_3 C_3)\right] \\
A_2 &= (L_2 - 1)\left(\beta_{2,2} + \beta_{2,3}\right) \\
C_t &= (L_2 - 1)F_t\left(1 - \hat{Q}_{t,2} + \beta_{t,2}\right)^2.
\end{aligned}$$

Illustration of the approximation process



Find Approximation

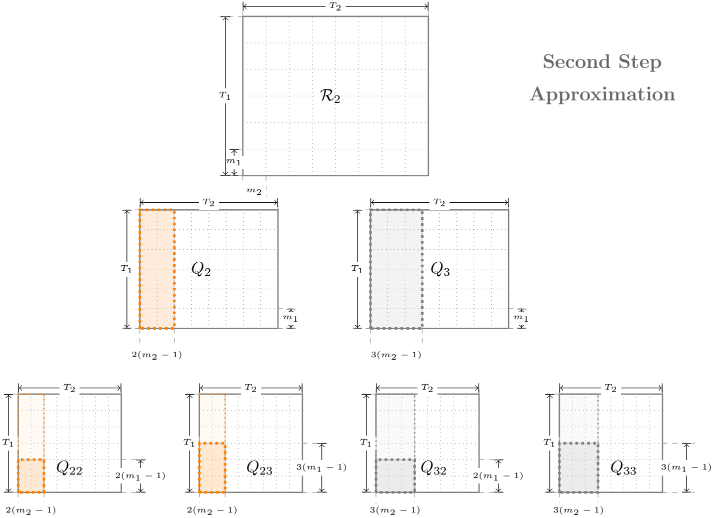
Illustration of the approximation process



First Step

Approximation

Illustration of the approximation process



Second Step

Approximation

Outline

Simulation methods for Normal data

Importance sampling algorithm

Test the null hypothesis of randomness against an alternative of clustering

H_0: The r.v.'s X_{s_1,s_2} are i.i.d. $\mathcal{N}(\mu,\sigma^2)$

H_1: There exists $\mathcal{R}(i_1,i_2)=[i_1-1,i_1+m_1-1]\times[i_2-1,i_2+m_2-1]\subset\mathcal{R}_2$ where the r.v.'s $X_{s_1,s_2}\sim\mathcal{N}(\mu_1,\sigma^2)$, $\mu_1>\mu$ and $X_{s_1,s_2}\sim\mathcal{N}(\mu,\sigma^2)$ outside $\mathcal{R}(i_1,i_2)$

Objective

Find a good estimate for $\mathbb{P}_{H_0}\left(S_{\mathbf{m}}(\mathbf{T};\mathcal{S})\geq\tau\right)$.

We are interested in evaluating the probability

$$\mathbb{P}_{H_0}\left(S_{\mathbf{m}}(\mathbf{T};\mathcal{S})\geq\tau\right)=\mathbb{P}\left(\bigcup_{i_1=1}^{T_1-m_1+1}\bigcup_{i_2=1}^{T_2-m_2+1}E_{i_1,i_2}(\mathcal{S})\right)$$

where $E_{i_1,i_2}(\mathcal{S})=\{Y_{i_1,i_2}(\mathcal{S})\geq\tau\}$.

Importance sampling algorithm

Algorithm 1 Importance Sampling Algorithm for Scan Statistics

Begin
Repeat for each k from 1 to $ITER$ (iterations number)

1: Generate uniformly the couple $(i_1^{(k)}, i_2^{(k)})$ from the set $\{1, \dots, T_1 - m_1 + 1\} \times \{1, \dots, T_2 - m_2 + 1\}$.

2: Given the couple $(i_1^{(k)}, i_2^{(k)})$, generate a sample of the random field $\tilde{\mathbf{X}}^{(k)} = \left\{\tilde{X}_{s_1, s_2}^{(k)}\right\}$, with $s_j \in \{1, \dots, T_j\}$ and $j \in \{1, 2\}$, from the conditional distribution of $\mathbf{X}$ given $\left\{Y_{i_1^{(k)}, i_2^{(k)}}(\mathcal{S}) \geq \tau\right\}$.

3: Take $c_k = C(\tilde{\mathbf{X}}^{(k)})$ the number of all couples (i_1, i_2) for which $\tilde{Y}_{i_1, i_2}(\mathcal{S}) \geq \tau$ and put $\widehat{\rho}_k(2) = \frac{1}{c_k}$.

End Repeat
Return

$$\widehat{\rho}(2) = \frac{1}{ITER} \sum_{k=1}^{ITER} \widehat{\rho}_k(2), \quad Var\left[\widehat{\rho}(2)\right] \approx \frac{1}{ITER - 1} \sum_{k=1}^{ITER} \left(\widehat{\rho}_k(2) - \frac{1}{ITER} \sum_{k=1}^{ITER} \widehat{\rho}_k(2)\right)^2$$

End

Importance sampling algorithm: $\mathcal{N}(\mu, \sigma^2)$

Step 2 requires to sample:

- $Y_{i_1^{(k)}, i_2^{(k)}}(\mathcal{S})$ from the tail distribution $\mathbb{P}\left(Y_{i_1^{(k)}, i_2^{(k)}}(\mathcal{S}) \geq \tau\right)$ ([Devroye, 1986])
- for the other indices, from the conditional distribution given $\left\{Y_{i_1^{(k)}, i_2^{(k)}}(\mathcal{S}) \geq \tau\right\}$

Lemma (Generalization of [Amărioarei, 2014, Lemma 3.4.4])

Let N be a positive integer, $\mathbf{X} = (X_1, X_2, \dots, X_N)$ be a vector of i.i.d. $\mathcal{N}(\mu, \sigma^2)$ and $\mathbf{a} = (a_1, \dots, a_N) \in \mathbb{R}^N$ a non zero constant vector ($a_j \neq 0$ for some particular j). Then conditionally given $\langle \mathbf{a}, \mathbf{X} \rangle = t$, the r.v.'s X_s with $s \neq j$ are jointly distributed as

$$\tilde{X}_s = \frac{a_s}{\|a\|}\left[\frac{t-\mu a_j}{\|a\|} - \frac{1}{\|a\|-|a_j|}\sum_{i\neq j} a_i\left(Z_i - \frac{\mu|a_j|}{\|a\|}\right)\right] + Z_s$$

where Z_s are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ r.v.s.

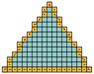
Outline

Numerical examples

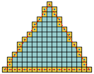
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 1 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Triangle

Window's shape — Triangle ($m_1 = 14$, $m_2 = 18$, $Nt = 133$, $IS = 1e4$, $IA = 1e5$)



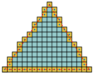
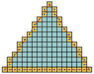
$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.916175	0.919820	0.010823	59	0.862501	0.894851	0.012021
4	0.997471	0.997507	0.000257	60	0.935163	0.948980	0.005516
5	0.999946	0.999946	0.000005	61	0.972665	0.975087	0.002637
6	0.999999	0.999999	0	62	0.986982	0.988558	0.001273
7	0.999999	0.999999	0	63	0.994493	0.994676	0.000599
8	1.000000	1.000000	0	64	0.997532	0.997604	0.000280
9	1.000000	1.000000	0	65	0.998943	0.998978	0.000126
10	1.000000	1.000000	0	66	0.999556	0.999562	0.000057
11	1.000000	1.000000	0	67	0.999824	0.999819	0.000025



$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.737933	0.822902	0.024424	50	0.912185	0.932616	0.007379
60	0.871787	0.903293	0.011103	51	0.946149	0.957850	0.004762
61	0.938941	0.951589	0.005389	52	0.966194	0.974052	0.003052
62	0.972291	0.975560	0.002639	53	0.980486	0.982878	0.001962
63	0.987388	0.988468	0.001309	54	0.988291	0.989444	0.001243
64	0.994128	0.994332	0.000642	55	0.993024	0.993292	0.000805
65	0.997330	0.997386	0.000307	56	0.995643	0.996031	0.000505
66	0.998794	0.998844	0.000147	57	0.997513	0.997494	0.000318
67	0.999474	0.999490	0.000068	58	0.998482	0.998597	0.000198

Scanning a region of size $T_1 \times T_2 = 250 \times 250$

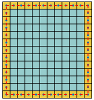
Table 1 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Triangle

Window's shape	Triangle ($m_1 = 14,\ m_2 = 18,\ Nt = 133,\ IS = 1e4,\ IA = 1e5$)							
	$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	3	0.916175	0.919820	0.010823	59	0.862501	0.894851	0.012021
	4	0.997471	0.997507	0.000257	60	0.935163	0.948980	0.005516
	5	0.999946	0.999946	0.000005	61	0.972665	0.975087	0.002637
	6	0.999999	0.999999	0	62	0.986982	0.988558	0.001273
	7	0.999999	0.999999	0	63	0.994493	0.994676	0.000599
	8	1.000000	1.000000	0	64	0.997532	0.997604	0.000280
	9	1.000000	1.000000	0	65	0.998943	0.998978	0.000126
	10	1.000000	1.000000	0	66	0.999556	0.999562	0.000057
	11	1.000000	1.000000	0	67	0.999824	0.999819	0.000025
	$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
	τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
	59	0.737933	0.822902	0.024424	50	0.912185	0.932616	0.007379
	60	0.871787	0.903293	0.011103	51	0.946149	0.957850	0.004762
	61	0.938941	0.951589	0.005389	52	0.966194	0.974052	0.003052
	62	0.972291	0.975560	0.002639	53	0.980486	0.982878	0.001962
	63	0.987388	0.988468	0.001309	54	0.988291	0.989444	0.001243
	64	0.994128	0.994332	0.000642	55	0.993024	0.993292	0.000805
	65	0.997330	0.997386	0.000307	56	0.995643	0.996031	0.000505
	66	0.998794	0.998844	0.000147	57	0.997513	0.997494	0.000318
	67	0.999474	0.999490	0.000068	58	0.998482	0.998597	0.000198

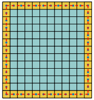
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 2 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Rectangle

Window's shape	Rectangle $(m_1 = 11,\ m_2 = 12,\ Nt = 132,\ IS = 1e4,\ IA = 1e5)$



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.947673	0.947757	0.005972	59	0.857404	0.856832	0.012641
4	0.997963	0.997979	0.000180	60	0.920590	0.919603	0.005960
5	0.999942	0.999942	0.000004	61	0.956954	0.956530	0.002886
6	0.999998	0.999998	0	62	0.977155	0.977179	0.001395
7	0.999999	0.999999	0	63	0.988161	0.988263	0.000669
8	1.000000	1.000000	0	64	0.994206	0.994089	0.000319
9	1.000000	1.000000	0	65	0.997114	0.997114	0.000149
10	1.000000	1.000000	0	66	0.998593	0.998613	0.000069
11	1.000000	1.000000	0	67	0.999341	0.999819	0.000031

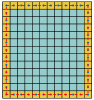


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.763228	0.763426	0.025991	50	0.864066	0.866081	0.010457
60	0.856115	0.858712	0.012264	51	0.904818	0.903116	0.006651
61	0.918130	0.918395	0.005941	52	0.932521	0.932663	0.004281
62	0.954831	0.954517	0.002958	53	0.954031	0.953874	0.002751
63	0.975084	0.975426	0.001483	54	0.968702	0.968700	0.001775
64	0.986979	0.987017	0.000735	55	0.978715	0.978566	0.001145
65	0.993240	0.993240	0.000361	56	0.985771	0.985812	0.000731
66	0.996511	0.996547	0.000176	57	0.990663	0.990575	0.000468
67	0.998276	0.998279	0.000084	58	0.993763	0.993827	0.000296

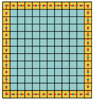
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 2 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Rectangle

Window's shape: Rectangle ($m_1 = 11,\ m_2 = 12,\ Nt = 132,\ IS = 1e4,\ IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.947673	0.947757	0.005972	59	0.857404	0.856832	0.012641
4	0.997963	0.997979	0.000180	60	0.920590	0.919603	0.005960
5	0.999942	0.999942	0.000004	61	0.956954	0.956530	0.002886
6	0.999998	0.999998	0	62	0.977155	0.977179	0.001395
7	0.999999	0.999999	0	63	0.988161	0.988263	0.000669
8	1.000000	1.000000	0	64	0.994206	0.994089	0.000319
9	1.000000	1.000000	0	65	0.997114	0.997114	0.000149
10	1.000000	1.000000	0	66	0.998593	0.998613	0.000069
11	1.000000	1.000000	0	67	0.999341	0.999819	0.000031

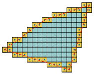


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.763228	0.763426	0.025991	50	0.864066	0.866081	0.010457
60	0.856115	0.858712	0.012264	51	0.904818	0.903116	0.006651
61	0.918130	0.918395	0.005941	52	0.932521	0.932663	0.004281
62	0.954831	0.954517	0.002958	53	0.954031	0.953874	0.002751
63	0.975084	0.975426	0.001483	54	0.968702	0.968700	0.001775
64	0.986979	0.987017	0.000735	55	0.978715	0.978566	0.001145
65	0.993240	0.993240	0.000361	56	0.985771	0.985812	0.000731
66	0.996511	0.996547	0.000176	57	0.990663	0.990575	0.000468
67	0.998276	0.998279	0.000084	58	0.993763	0.993827	0.000296

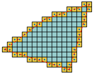
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 3 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Quadrilateral

Window's shape — **Quadrilateral** ($m_1 = 14,\ m_2 = 18,\ Nt = 131,\ IS = 1e4,\ IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.925258	0.926174	0.009722	59	0.913203	0.926709	0.007540
4	0.997615	0.997619	0.000235	60	0.958806	0.963738	0.003582
5	0.999945	0.999945	0.000005	61	0.981433	0.982643	0.001725
6	0.999999	0.999999	0	62	0.991635	0.991891	0.000823
7	0.999999	0.999999	0	63	0.996170	0.996220	0.000387
8	1.000000	1.000000	0	64	0.998287	0.998322	0.000180
9	1.000000	1.000000	0	65	0.999243	0.999240	0.000082
10	1.000000	1.000000	0	66	0.999679	0.999693	0.000037
11	1.000000	1.000000	0	67	0.999866	0.999869	0.000016

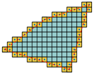


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.836771	0.871611	0.015388	50	0.920571	0.935366	0.006875
60	0.918389	0.931407	0.007235	51	0.949483	0.957631	0.004378
61	0.960554	0.963979	0.003555	52	0.968977	0.972699	0.002844
62	0.981340	0.982444	0.001754	53	0.980772	0.982865	0.001823
63	0.991148	0.991396	0.000863	54	0.988182	0.989097	0.001175
64	0.995798	0.996026	0.000423	55	0.992870	0.992982	0.000758
65	0.998148	0.998168	0.000205	56	0.995739	0.995736	0.000481
66	0.999126	0.999157	0.000097	57	0.997355	0.997429	0.000302
67	0.999618	0.999621	0.000045	58	0.998397	0.998445	0.000190

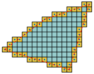
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 3 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Quadrilateral

Window's shape — **Quadrilateral ($m_1 = 14,\ m_2 = 18,\ Nt = 131,\ IS = 1e4,\ IA = 1e5$)**



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.925258	0.926174	0.009722	59	0.913203	0.926709	0.007540
4	0.997615	0.997619	0.000235	60	0.958806	0.963738	0.003582
5	0.999945	0.999945	0.000005	61	0.981433	0.982643	0.001725
6	0.999999	0.999999	0	62	0.991635	0.991891	0.000823
7	0.999999	0.999999	0	63	0.996170	0.996220	0.000387
8	1.000000	1.000000	0	64	0.998287	0.998322	0.000180
9	1.000000	1.000000	0	65	0.999243	0.999240	0.000082
10	1.000000	1.000000	0	66	0.999679	0.999693	0.000037
11	1.000000	1.000000	0	67	0.999866	0.999869	0.000016

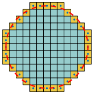


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.836771	0.871611	0.015388	50	0.920571	0.935366	0.006875
60	0.918389	0.931407	0.007235	51	0.949483	0.957631	0.004378
61	0.960554	0.963979	0.003555	52	0.968977	0.972699	0.002844
62	0.981340	0.982444	0.001754	53	0.980772	0.982865	0.001823
63	0.991148	0.991396	0.000863	54	0.988182	0.989097	0.001175
64	0.995798	0.996026	0.000423	55	0.992870	0.992982	0.000758
65	0.998148	0.998168	0.000205	56	0.995739	0.995736	0.000481
66	0.999126	0.999157	0.000097	57	0.997355	0.997429	0.000302
67	0.999618	0.999621	0.000045	58	0.998397	0.998445	0.000190

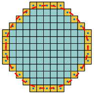
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 4 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Circle

Window's shape: Circle ($m_1 = 13,\ m_2 = 13,\ Nt = 129,\ IS = 1e4,\ IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.949902	0.950399	0.006284	59	0.920134	0.920987	0.005962
4	0.998131	0.998099	0.000187	60	0.956871	0.957137	0.002840
5	0.999947	0.999947	0.000004	61	0.977475	0.977539	0.001366
6	0.999998	0.999998	0	62	0.988597	0.988595	0.000651
7	0.999999	0.999999	0	63	0.994265	0.994274	0.000306
8	1.000000	1.000000	0	64	0.997227	0.997257	0.000142
9	1.000000	1.000000	0	65	0.998698	0.998675	0.000065
10	1.000000	1.000000	0	66	0.999377	0.999381	0.000029
11	1.000000	1.000000	0	67	0.999709	0.999715	0.000013

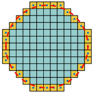


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.860740	0.860278	0.012220	50	0.889276	0.889385	0.008241
60	0.918923	0.919454	0.005955	51	0.920701	0.920950	0.005242
61	0.955942	0.955339	0.002915	52	0.945380	0.945514	0.003351
62	0.975615	0.975720	0.001449	53	0.962727	0.962884	0.002142
63	0.987403	0.987314	0.000712	54	0.974988	0.974968	0.001370
64	0.993559	0.993485	0.000347	55	0.983200	0.983188	0.000875
65	0.996698	0.996696	0.000167	56	0.989078	0.988928	0.000556
66	0.998365	0.998353	0.000079	57	0.992649	0.992725	0.000350
67	0.999213	0.999202	0.000037	58	0.995300	0.995278	0.000220

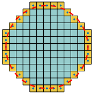
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 4 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Circle

Window's shape — Circle ($m_1 = 13,\ m_2 = 13,\ Nt = 129,\ IS = 1e4,\ IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.949902	0.950399	0.006284	59	0.920134	0.920987	0.005962
4	0.998131	0.998099	0.000187	60	0.956871	0.957137	0.002840
5	0.999947	0.999947	0.000004	61	0.977475	0.977539	0.001366
6	0.999998	0.999998	0	62	0.988597	0.988595	0.000651
7	0.999999	0.999999	0	63	0.994265	0.994274	0.000306
8	1.000000	1.000000	0	64	0.997227	0.997257	0.000142
9	1.000000	1.000000	0	65	0.998698	0.998675	0.000065
10	1.000000	1.000000	0	66	0.999377	0.999381	0.000029
11	1.000000	1.000000	0	67	0.999709	0.999715	0.000013

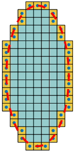


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.860740	0.860278	0.012220	50	0.889276	0.889385	0.008241
60	0.918923	0.919454	0.005955	51	0.920701	0.920950	0.005242
61	0.955942	0.955339	0.002915	52	0.945380	0.945514	0.003351
62	0.975615	0.975720	0.001449	53	0.962727	0.962884	0.002142
63	0.987403	0.987314	0.000712	54	0.974988	0.974968	0.001370
64	0.993559	0.993485	0.000347	55	0.983200	0.983188	0.000875
65	0.996698	0.996696	0.000167	56	0.989078	0.988928	0.000556
66	0.998365	0.998353	0.000079	57	0.992649	0.992725	0.000350
67	0.999213	0.999202	0.000037	58	0.995300	0.995278	0.000220

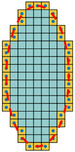
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 5 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Ellipse

Window's shape	Ellipse ($m_1 = 19$, $m_2 = 9$, $Nt = 135$, $IS = 1e4$, $IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.945149	0.944018	0.006856	59	0.761436	0.764423	0.022290
4	0.997692	0.997783	0.000217	60	0.861070	0.858224	0.011120
5	0.999934	0.999935	0.000005	61	0.922564	0.920545	0.005593
6	0.999998	0.999998	0	62	0.956994	0.956925	0.002803
7	0.999999	0.999999	0	63	0.977214	0.977111	0.001391
8	1.000000	1.000000	0	64	0.988164	0.988170	0.000678
9	1.000000	1.000000	0	65	0.993942	0.994013	0.000328
10	1.000000	1.000000	0	66	0.997061	0.997005	0.000155
11	1.000000	1.000000	0	67	0.998561	0.998565	0.000072

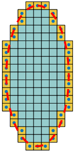


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.638328	0.638600	0.042412	50	0.845330	0.842760	0.011460
60	0.769531	0.769090	0.021264	51	0.887757	0.887823	0.007568
61	0.860703	0.859829	0.010885	52	0.921023	0.920804	0.005026
62	0.918782	0.919289	0.005606	53	0.944751	0.944745	0.003312
63	0.954499	0.954971	0.002890	54	0.961667	0.961830	0.002176
64	0.975360	0.975217	0.001468	55	0.973786	0.973964	0.001431
65	0.986788	0.986846	0.000746	56	0.982379	0.982058	0.000929
66	0.993076	0.993056	0.000372	57	0.988009	0.988142	0.000599
67	0.996464	0.996456	0.000183	58	0.992139	0.992160	0.000385

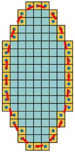
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 5 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Ellipse

Window's shape	Ellipse ($m_1 = 19$, $m_2 = 9$, $Nt = 135$, $IS = 1e4$, $IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.945149	0.944018	0.006856	59	0.761436	0.764423	0.022290
4	0.997692	0.997783	0.000217	60	0.861070	0.858224	0.011120
5	0.999934	0.999935	0.000005	61	0.922564	0.920545	0.005593
6	0.999998	0.999998	0	62	0.956994	0.956925	0.002803
7	0.999999	0.999999	0	63	0.977214	0.977111	0.001391
8	1.000000	1.000000	0	64	0.988164	0.988170	0.000678
9	1.000000	1.000000	0	65	0.993942	0.994013	0.000328
10	1.000000	1.000000	0	66	0.997061	0.997005	0.000155
11	1.000000	1.000000	0	67	0.998561	0.998565	0.000072

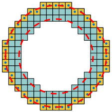


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.638328	0.638600	0.042412	50	0.845330	0.842760	0.011460
60	0.769531	0.769090	0.021264	51	0.887757	0.887823	0.007568
61	0.860703	0.859829	0.010885	52	0.921023	0.920804	0.005026
62	0.918782	0.919289	0.005606	53	0.944751	0.944745	0.003312
63	0.954499	0.954971	0.002890	54	0.961667	0.961830	0.002176
64	0.975360	0.975217	0.001468	55	0.973786	0.973964	0.001431
65	0.986788	0.986846	0.000746	56	0.982379	0.982058	0.000929
66	0.993076	0.993056	0.000372	57	0.988009	0.988142	0.000599
67	0.996464	0.996456	0.000183	58	0.992139	0.992160	0.000385

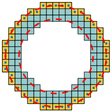
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 6 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Annulus

Window's shape — **Annulus ($m_1 = 17$, $m_2 = 17$, $Nt = 124$, $IS = 1e4$, $IA = 1e5$)**



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.881687	0.882372	0.009897	59	0.951321	0.951283	0.001572
4	0.995456	0.995488	0.000210	60	0.975645	0.975784	0.000654
5	0.999882	0.999883	0.000004	61	0.988191	0.988284	0.000280
6	0.999997	0.999997	0	62	0.994451	0.994451	0.000120
7	0.999999	0.999999	0	63	0.997437	0.997444	0.000051
8	1.000000	1.000000	0	64	0.998835	0.998838	0.000021
9	1.000000	1.000000	0	65	0.99948	0.999483	0.000009
10	1.000000	1.000000	0	66	0.999774	0.999774	0.000003
11	1.000000	1.000000	0	67	0.999903	0.999903	0.000001

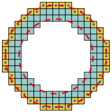


$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.902926	0.903783	0.003972	50	0.860688	0.860266	0.006625
60	0.948792	0.949016	0.001623	51	0.905117	0.904708	0.003571
61	0.973906	0.973904	0.000700	52	0.936387	0.935930	0.001991
62	0.986863	0.986917	0.000311	53	0.957907	0.957630	0.001137
63	0.993560	0.993580	0.000138	54	0.972310	0.972343	0.000661
64	0.996904	0.996901	0.000061	55	0.982134	0.982127	0.000386
65	0.998536	0.998536	0.000027	56	0.988540	0.988555	0.000228
66	0.999318	0.999318	0.000011	57	0.992783	0.992775	0.000135
67	0.999688	0.999689	0.000005	58	0.995458	0.995467	0.000079

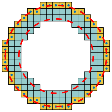
Scanning a region of size $T_1 \times T_2 = 250 \times 250$

Table 6 : Numerical results for $\mathbb{P}(S \leqslant \tau)$: Annulus

Window's shape	Annulus ($m_1 = 17$, $m_2 = 17$, $Nt = 124$, $IS = 1e4$, $IA = 1e5$)



$X_{s_1,s_2} \sim \mathcal{B}(1, 0.01)$				$X_{s_1,s_2} \sim \mathcal{B}(5, 0.05)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
3	0.881687	0.882372	0.009897	59	0.951321	0.951283	0.001572
4	0.995456	0.995488	0.000210	60	0.975645	0.975784	0.000654
5	0.999882	0.999883	0.000004	61	0.988191	0.988284	0.000280
6	0.999997	0.999997	0	62	0.994451	0.994451	0.000120
7	0.999999	0.999999	0	63	0.997437	0.997444	0.000051
8	1.000000	1.000000	0	64	0.998835	0.998838	0.000021
9	1.000000	1.000000	0	65	0.99948	0.999483	0.000009
10	1.000000	1.000000	0	66	0.999774	0.999774	0.000003
11	1.000000	1.000000	0	67	0.999903	0.999903	0.000001



$X_{s_1,s_2} \sim \mathcal{P}(0.25)$				$X_{s_1,s_2} \sim \mathcal{N}(0, 1)$			
τ	Sim	AppH	ETotal	τ	Sim	AppH	ETotal
59	0.902926	0.903783	0.003972	50	0.860688	0.860266	0.006625
60	0.948792	0.949016	0.001623	51	0.905117	0.904708	0.003571
61	0.973906	0.973904	0.000700	52	0.936387	0.935930	0.001991
62	0.986863	0.986917	0.000311	53	0.957907	0.957630	0.001137
63	0.993560	0.993580	0.000138	54	0.972310	0.972343	0.000661
64	0.996904	0.996901	0.000061	55	0.982134	0.982127	0.000386
65	0.998536	0.998536	0.000027	56	0.988540	0.988555	0.000228
66	0.999318	0.999318	0.000011	57	0.992783	0.992775	0.000135
67	0.999688	0.999689	0.000005	58	0.995458	0.995467	0.000079

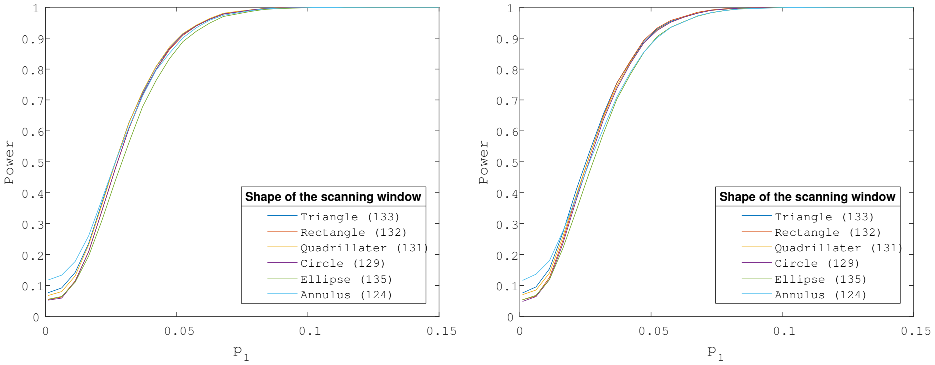
Outline

Power of the scan statistic test

Power evaluation for $\mathcal{B}(1, 0.001)$ model

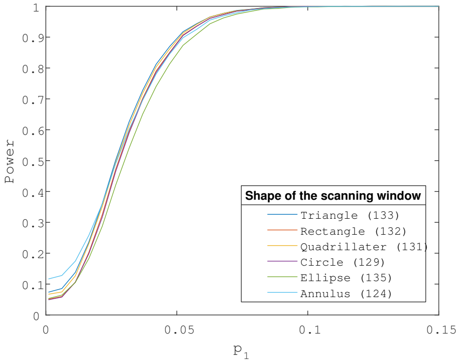
Triangular simulated cluster

Rectangular simulated cluster

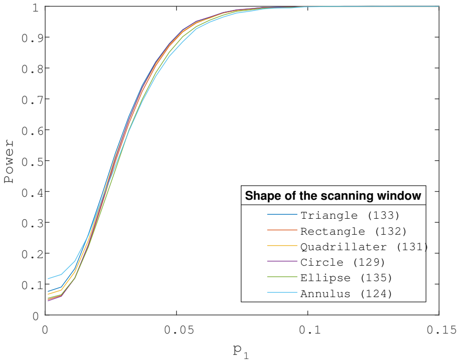


Power evaluation for $\mathcal{B}(1, 0.001)$ model

Quadrilateral simulated cluster

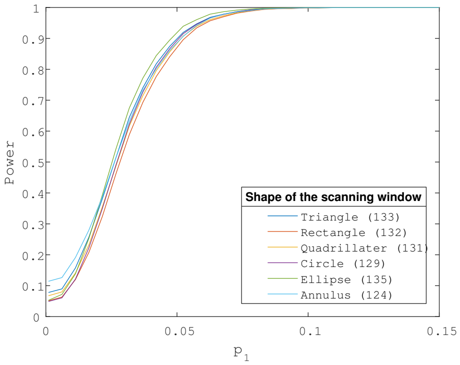


Circular simulated cluster

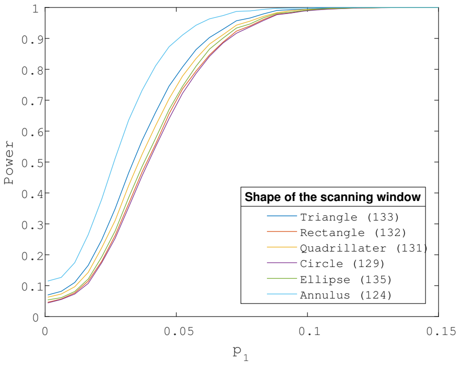


Power evaluation for $\mathcal{B}(1, 0.001)$ model

Ellipsoidal simulated cluster



Annular simulated cluster




thank you!

Alm, S. E. (1983).
On the distribution of scan statistic of poisson process.
In Gut, A. and Helst, L., editors, *Probability and Mathematical Statistics*, pages 1–10. Upsalla University Press.

Alm, S. E. (1997).
On the distributions of scan statistics of a two-dimensional poisson process.
Advances in Applied Probability, pages 1–18.

Alm, S. E. (1998).
Approximation and simulation of the distributions of scan statistics for poisson processes in higher dimensions.
Extremes, 1(1):111–126.

Amărioarei, A. (2014).
Approximations for the multidimensional discrete scan statistics.
PhD thesis, University of Lille 1.

Anderson, N. H. and Titterington, D. M. (1997).
Some methods for investigating spatial clustering, with epidemiological applications.
Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(1):87–105.

Assunção, R., Costa, M., Tavares, A., and Ferreira, S. (2006).
Fast detection of arbitrarily shaped disease clusters.
Statistics in Medicine, 25(5):723–742.

Bresenham, J. (1965).
Algorithm for computer control of a digital printer.
IBM Systems Journal, 4(1).

Bresenham, J. (1977).
A linear algorithm for incremental digital display of circular arcs.
Communications of the ACM, 20(2):100–106.

Devroye, L. (1986).
Non uniform random variate generation.
Springer-Verlag, New York.

Foley, J. (1995).
Computer Graphics, Principles and Practice in C.
Addison-Wesley Professional.

Glaz, J., Naus, J., and Wallenstein, S. (2001).
Scan Statistics.

Springer.

Glaz, J., Pozdnyakov, V., and Wallenstein, S. (2009).
Scan Statistics: Methods and Applications.
Birkhaüser Boston.

Loader, C. R. (1991).
Large-deviation approximations to the distribution of scan statistics.
Advances in Applied Probability, pages 751–771.

Tango, T. and Takahashi, K. (2005).
A flexibly shaped spatial scan statistic for detecting clusters.
Int J Health Geogr, 4:11.
Tango, Toshiro Takahashi, Kunihiko England Int J Health Geogr. 2005 May 18;4:11.