

# Preface

This master thesis, written during the spring of 2013 at the Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, marks the end of my five year study program, Environmental Physics and Renewable Energy. The aim of this thesis has been to develop statistical test for some of the main connection algorithms used in neural networks.

The subject of this thesis is located somewhere in the intersection between the fields of statistics, computer science and neuroscience. I have some basic training in the first two, but the last one was completely new to me before I started this work. That was partly why I chose this subject. It was an opportunity to gain some insight into this exciting field, that is unraveling the mysteries of the human mind, and might come to shape technology in the future. Working on this project has been just as educational and interesting as I had hoped for.

I would like to thank my supervisor, Hans Ekkehard Plessner, and my co-supervisor, Birgit Kriener, for their guidance along the way, and for always taking the time to answer my questions thoroughly. Our fruitful conversations have been of great help. I would also like to thank my parents for their unconditional love and support throughout my life, and for cheering for me from the sidelines through this project. Last, but not least, I would like to thank Elina. You have been very supportive, and always patient, even through the most stressful times, when I was glued to my computer for hours on end, unable to tear myself away.



# Abstract

Simulations of increasingly sophisticated neural network models have accelerated the progress of neuroscience, but have also led to an increased reliance on the simulation software. Testing the quality of this software therefore becomes important. Here we develop test strategies for some of the most common probabilistic connection algorithms used in simulators, and implement these as Python based test suites. We develop approaches to alleviate the problems of statistical software testing, i.e., the unavoidable occurrence of false positives and negatives. The tests are developed for the NEST simulator, but can easily be adapted to work with analogous connection algorithms in other simulators.

For random connections with predefined in- or out-degree, observed random degrees are compared with expectation using Pearson's chi-squared test. For networks with structure in two- or three-dimensional space, i.e., with a distance-dependent connection probability, the Kolmogorov-Smirnov (KS) test is used to compare the empirical distribution function (EDF) of distances between connected source-target pairs with the expected cumulative distribution (CDF), obtained by numerical integration of the normalized product of the radial distribution of nodes and the distance-dependent connection probability (kernel). A Z-test comparing the total number of connections with the expectation is also implemented. For all three of these tests, a two-level test can be used, comparing the distribution of  $p$ -values from multiple tests of individual network realizations with the expected uniform distribution, using the KS test. This approach results in greatly increased sensitivity. For automated tests used in test suites, an adaptive approach is proposed. Here, one test is performed, and if the result is suspicious, the more thorough two-level test is performed. This approach is fast, as the two-level test is only invoked for a fraction of the test cases under normal circumstances. It also results in a very low rate of false positives.

The probabilistic connection algorithms of NEST were tested under a variety of conditions, e.g. with different network sizes, different number of virtual processes (VPs) and different distance-dependent connection probability functions (kernels). No evidence of any error or bias was found in the algorithms. The test strategies themselves were shown to detect an array of small errors and biases when these were deliberately introduced into the algorithm.



# Sammendrag

Simuleringer av stadig mer sofistikerte nevrale nettverksmodeller har vært med på å drive nevrovitenskapen fremover, men har samtidig ført til en økt avhengighet av simuleringsprogramvare. Å teste kvaliteten til denne programvaren derfor viktig. Her utvikler vi strategier for testing av noen av de mest brukte koblingsalgoritmene, og vi implementerer disse som Python-baserte testpakker. Vi utvikler metoder for å redusere hyppigheten av type I og type II feil. Testene er utviklet for simulatoren NEST, men kan modifiseres for å fungere med tilsvarende koblingsalgoritmer i andre simulatorer.

For tilfeldige koblinger med forhåndsbestemt inn- eller utgrad sammenlignes de observerte tilfeldige gradene med forventningsverdier ved hjelp av Pearsons kjikvadrattest. For nettverk med romlig struktur, dvs. nettverk med en koblingssannsynlighet som avhenger av avstand, brukes Kolmogorov-Smirnov-testen (KS-testen) til å sammenligne den empiriske kumulative fordelingen av avstander mellom sammenkoblede node-par med den forventede kumulative fordelingen, funnet ved numerisk integrasjon av det normaliserte produktet av den radiale fordelingsfunksjonen av noder og den avstandsavhengige koblingssannsynligheten. En Z-test som sammenligner det totale antallet koblinger med forventningsverdien er også implementert. For alle disse testene kan en to-nivå-test anvendes. Denne sammenligner fordelingen av  $p$ -verdier fra flere tester av individuelle nettverksrealiseringer med den forventede uniforme fordelingen ved hjelp av KS-testen. Dette gir en betraktelig økt sensitivitet. For automatiserte tester brukt i testpakker foreslår vi en adaptiv løsning. Denne går ut på at en enkel test kjøres, og kun dersom resultatet er mistenklig kjøres den mer grundige to-nivå-testen. På denne måten kan automatiserte tester kjøre fort, siden to-nivå-testen normalt kun vil kjøres for en liten andel av test-tilfellene. I tillegg får vi få feilaktige godkjennelser (type I feil).

De tilfeldige koblingsalgoritmene i NEST ble testet under ulike forhold, for eksempel med forskjellig nettverkstørrelse, forskjellig antall virtuelle prosesser og foreskjellige avstandsavhengige koblingssannsynligheter. Det ble ikke funnet noe bevis på feil eller skjevheter i algoritmene. Det ble vist at teststrategiene oppdaget en rekke feil og skjevheter når disse bevisst ble lagt inn i algoritmene.



# Contents

<b>Preface</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Sammendrag</b>	<b>v</b>
<b>List of Figures</b>	<b>x</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Symbols</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Statistical testing</b>	<b>7</b>
2.1 Pearson's chi-squared test . . . . .	8
2.2 The Kolmogorov-Smirnov test . . . . .	11
2.3 The Z-test . . . . .	13
<b>3 Distribution of connections</b>	<b>15</b>
3.1 Random convergent connections . . . . .	15
3.1.1 Implementation . . . . .	16
3.1.2 Results . . . . .	17
3.2 Random divergent connections . . . . .	22
3.2.1 Implementation . . . . .	22
3.2.2 Results . . . . .	23
3.3 Automated test procedure . . . . .	23
3.3.1 Implementation . . . . .	24
<b>4 Spatially structured networks</b>	<b>27</b>
4.1 Two-dimensional space . . . . .	27
4.1.1 Implementation . . . . .	30
4.1.2 Results . . . . .	32
4.2 Three-dimensional space . . . . .	35

4.2.1	Implementation . . . . .	38
4.2.2	Results . . . . .	38
4.3	Automated test procedure . . . . .	40
4.3.1	Implementation . . . . .	41
<b>5</b>	<b>Discussion</b>	<b>43</b>
<b>A</b>	<b>NEST - A short tutorial</b>	<b>47</b>
<b>B</b>	<b>Test script for random convergent connections</b>	<b>59</b>
<b>C</b>	<b>Test script for random divergent connections</b>	<b>63</b>
<b>D</b>	<b>Automated test suite for random convergent and divergent connections</b>	<b>67</b>
<b>E</b>	<b>Test script for 2D spatially structured network</b>	<b>69</b>
<b>F</b>	<b>Test script for 3D spatially structured network</b>	<b>73</b>
<b>G</b>	<b>Automated test suite for spatially structured networks</b>	<b>77</b>
	<b>References</b>	<b>81</b>

# List of Figures

1.1	A random divergent connection pattern . . . . .	4
1.2	A spatially structured network . . . . .	5
2.1	CDF for $X \sim \mathcal{U}(1, 0)$ and EDF of pseudorandom numbers from a PRNG . . . . .	12
3.1	Histogram and EDF of $p$ -values from 10,000 chi-squared GOF tests . . . . .	18
3.2	EDF of 100 $p$ -values from the two-level tests procedure . . . . .	19
3.3	The effect of small parameter values on the EDF of $p$ -values . . . . .	20
3.4	Distribution of $p$ -values after introducing a bias into the connection algorithm . . . . .	21
4.1	Illustration of the derivation of the expression for the radial distribution function . . . . .	28
4.2	The effect of the mask on the radial distribution function . . . . .	29
4.3	Exemplary PDF, CDF and connectivity pattern for a network in 2D space, using a Gaussian kernel . . . . .	31
4.4	Theoretical and empirical PDF and CDF of source-target distances using a constant kernel . . . . .	32
4.5	Connectivity pattern with periodic boundary conditions and a source node shifted away from the layer's center . . . . .	33
4.6	Detection rate as a function of network size . . . . .	34
4.7	Illustration of the effect of a cubic mask on the radial distribution function . . . . .	36
4.8	Illustration of the derivation of the surface area $E$ of the intersection between adjacent spherical caps . . . . .	37
4.9	Exemplary PDF, CDF and connectivity pattern for a network in 3D space, using a Gaussian kernel . . . . .	38
4.10	Theoretical and empirical PDF and CDF of source-target distances using a constant kernel . . . . .	39
A.1	Example of free two-dimensional layer . . . . .	53
A.2	Example of free three-dimensional layer . . . . .	54

A.3 Example of connectivity pattern . . . . .	56
A.4 Distribution of distances between source node and connected target nodes . . . . .	57

# List of Tables

3.1	Default parameter set for reported results . . . . .	17
3.2	Default parameter set for sensitivity testing . . . . .	21
A.1	Neuron models in NEST . . . . .	48
A.2	Devices in NEST . . . . .	49
A.3	Synapse models in NEST . . . . .	50
A.4	Kernels in NEST Topology . . . . .	55



# List of Symbols

$\alpha$	level of significance / probability of making a type I error	p. 7
$\beta$	probability of making a type II error	p. 7
$\mathcal{U}(a, b)$	uniform distribution on the interval $[a, b]$	p. 7
$A_i$	outcome of a trial with a fixed number $r$ of possible outcomes	p. 9
$n$	number of trials	p. 9
$r$	number of possible outcomes $A_i$	p. 9
$p_i$	probability of outcome $A_i$	p. 9
$\mathbf{p}$	vector of outcome probabilities $p_i$	p. 9
$\nu_i$	frequency of outcome $A_i$	p. 9
$\boldsymbol{\nu}$	vector of frequencies	p. 9
$E(\nu_i)$	expected frequency of outcome $A_i$	p. 9
$E(\boldsymbol{\nu})$	vector of expected frequencies	p. 9
$X^2$	Pearson's test statistic	p. 10
$\chi^2$	the chi squared distribution	p. 10
$F(x)$	cumulative distribution function of a random variable	p. 11
$S_n(x)$	empirical distribution function that converges to $F(x)$	p. 11
$F_0(x)$	hypothesized cumulative distribution function	p. 11
$D_n$	Kolmogorov-Smirnov test statistic	p. 11
$Z$	standard score (z-score)	p. 13
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean $\mu$ and variance $\sigma^2$	p. 13
$C$	fixed in or out degree (number of connections) in network with random convergent or divergent connections, respectively	p. 15
$N_s$	number of source nodes in network	p. 15
$N_t$	number of target nodes in network	p. 15
$n_{\text{runs}}$	number of times to run chi-squared test in two-level test	p. 17
$L$	side length of square or cubic layer	p. 27
$N$	number of nodes in spatially structured network	p. 27

$D$	distance from center node	p. 27
$D_{ij}$	distance from node $i$ to node $j$	p. 27
$\rho_0$	areal or volumetric node density for networks in 2D or 3D space, respectively	p. 28
$\rho(D)$	radial distribution function (RDF)	p. 28
$\mathcal{P}(D)$	distance-dependent connection probability (kernel)	p. 29
$f(D)$	probability density function (PDF) of source-target distances	p. 29
$F(D)$	cumulative distribution function (CDF) of source-target distances	p. 29
$H(x)$	the Heaviside step function	p. 30
$C$	total number of connections in spatially structured network	p. 30

# Chapter 1

## Introduction

The brain is a truly remarkable machine. It routinely performs tasks that are impossible for even the most powerful supercomputer, and it does so in a highly energy-efficient manner. It can analyze patterns extremely efficiently, interpret them, and produce appropriate behavioral responses. It can store enormous amounts of information, allowing us to recollect things like phone numbers, familiar faces, songs and episodes from early life. But perhaps even more impressively, the brain is capable of producing consciousness, and allows us to experience our own thoughts and feelings.

Scientists' desire to unlock the mysteries of the brain is obvious. Not only would it allow us to better understand, and perhaps cure, the hundreds of different brain diseases that exist, such as depression, Alzheimer's disease, Parkinson's disease, epilepsy, and migraine. The knowledge of how the brain solves complex problems would likely also result in radical changes to how computers work. New storage technologies that mimic the way the brain stores and recollects memories might emerge, and computers might finally be able to learn the way humans do. Artificial intelligence (AI) might seem a lot less artificial.

However, understanding the brain is one of the greatest challenges facing scientists today. A great deal is understood about how single neurons function, how they communicate with other neurons through synapses, and how these synapses are formed and change over time. What is not well understood is how brain function emerges from billions of neurons communicating with each other over trillions of synapses. One way to address this problem is to simulate the activity of a large number of neurons in a computer. With such a simulation, the effects of small changes in neurons and synapses on the dynamics of large networks can be studied. The Human Brain Project (HBP), recently awarded one billion euros by the European Commission ([Abbott 2013](#)), aims to simulate a full scale human brain within a decade ([The Human Brain Project Preparatory Study Consortium 2012](#)). With this flagship project engaging and inspiring the scientific community, the importance of simulations in brain

science is likely to continue to increase in the years to come.

A number of neural network simulators are available, such as NEURON, GENESIS, Brian, and NEST ([Brette et al. 2007](#)). This thesis will focus on NEST ([Gewaltig and Diesmann 2007](#)), a popular simulation environment for simulating large networks of point neurons. As scientists are increasingly relying on tools like these to make new discoveries, the need for quality checking of the software increases. There is always a chance that a mistake has crept into a piece of computer code, one that does not make the program crash, but causes erroneous scientific results. In a famous example, a paper detailing the structure of a protein called MsbA ([Chang and Roth 2001](#)), had to be retracted because the reported structure turned out to be wrong. The reason for the incorrect result was that two columns of data were flipped in the computer program that derived the protein structure ([Miller 2006](#)). It took more than five years before the mistake was detected, and by that time the paper had been cited by 364 publications, according to Google Scholar. Four other papers also had to be retracted due to the same malfunctioning computer program.

A good way to detect these kinds of mistakes in computer code is through unit testing. The idea here is to divide the program into small units that can be tested rigorously by comparing the units output with our expectations ([Huizinga and Kolawa 2007](#)). While this approach has its merits, tests will not be able to detect all conceivable error that could occur. In the words of [Dijkstra et al. \(1970\)](#), “Program testing can be used to show the presence of bugs, but never to show their absence!”.

When testing probabilistic algorithms using statistical tests, some additional challenges arise. For instance, the tests will occasionally give false positives, i.e., they will report a problem with the tested algorithm where none exists. One can not get completely rid of false positives, but strategies can be devised to reduce the frequency with which they occur. Statistical tests will also give false negatives, i.e., they will fail to report true problems. Again, this is unavoidable, all we can do is to try to increase the sensitivity of the tests, i.e., their ability to detect problems. This ties in with a more general problem; that of testing randomness.

To generate a sequence of seemingly random numbers, computer software relies on pseudorandom number generators PRNGs ([L'Ecuyer 2004](#)). The sequences they produce is not truly random, but generated by a deterministic algorithm (hence the prefix “pseudo”). For most applications, however, the numbers are sufficiently unpredictable as to be considered random, depending on the specific algorithm used and how it is implemented. There are several ways to test the output of a PRNG. One approach is to look for a systematic bias by comparing the expected distribution of numbers with an observed sample distribution using some goodness-of-fit (GOF) test, such as the Kolmogorov-Smirnov (KS) test or the Anderson-Darling (AD) test, but this approach will not detect other kinds of patterns such as clustering of the

produced numbers, as long as the clusters are distributed relatively evenly. Other methods may pick up such clustering, but yet other unforeseeable patterns might still go undetected. As argued by [L'Ecuyer and Simard \(2007\)](#), no statistical test or battery of tests can guarantee that the output of a PRNG will be sufficiently random under all circumstances. Some application might cause some structure to emerge due to an artifact of the PRNG's algorithm.

When testing probabilistic algorithms using statistical tests, we are faced with the same problem. No statistical test or battery of tests can prove beyond any doubt that the algorithm is without biases or patterns that affects the output of a computer program. Still, the more tests the algorithms can pass, the greater our confidence in them, and hence the scientific findings based on them.

## Connectivity patterns in neuronal network models

A neural network can be described as a directed, weighted graph where the nodes are neurons, or sometimes devices, and the edges are connections or synapses between them. Events can be transmitted in one direction over the connections. These events are action potentials (spikes) or other types of signals that can be transmitted over synapses.

In any type of network modeling, the network structure must be specified. In principle, this could be done by listing all the nodes and their connections. This kind of specification, however, is not very manageable for us humans. To be able to work with network models, discuss them, and share them, higher-level descriptions of connection patterns are needed. Populations of nodes can then be connected following some basic rules to create these patterns. Ways to unambiguously describe and document connection patterns have been lacking, but recent efforts have been made to standardize their terminology and notation ([Nordlie and Plesser 2010](#); [Djurfeldt 2012](#); [Crook et al. 2012](#)). Since NEST is used in this thesis, we will mainly use NESTs terminology. We will now describe the relevant connection patterns and related concepts and terminology.

A **random divergent connection** between a source population and a target population is defined as the connection pattern resulting from connecting each node in the source population to a prescribed number  $C$  of randomly drawn target nodes. An example of this kind of network is shown in Figure 1.1. The target nodes are drawn with replacement, meaning multiple connections can exist between any pair of nodes (multapses). If the source and target populations are the same population, nodes will also be able to connect to themselves (autapses). It is possible to disallow multapses and autapses in NEST, but we will assume they are allowed.

In a **random convergent connection**,  $C$  *source nodes* are randomly drawn, with equal probability, for each target node. The in-degree is now  $C$

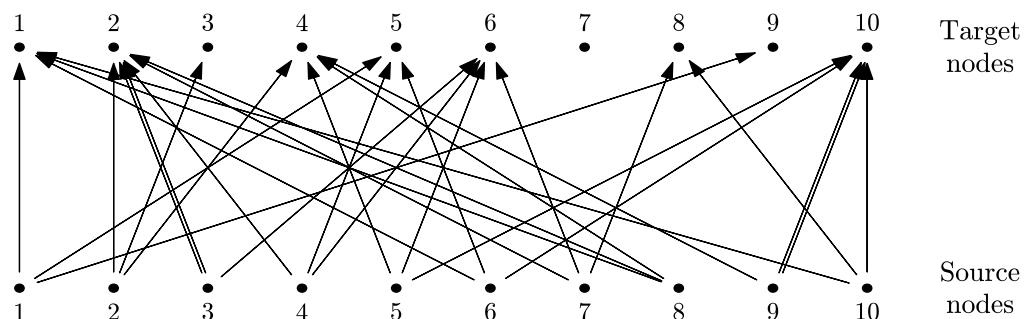
for all target nodes, while the source nodes might have different out-degrees.

Network models can also have spatial structure. In such models, connection probabilities, as well as connection properties such as weight and delay, can be a function of the distance between the source and the target node. In this way, connectivity patterns that mimic observed axonal projection patterns can be created. Thus, when referring to a **spatially structured network** in this thesis, we mean a network whose connection probabilities depend on source-target distance. In spatially structured networks, populations of nodes can exist in two- or three-dimensional space. The distance-dependent connection probability function is referred to as the **kernel**. Only nodes inside an area or volume called the **mask** are eligible connection targets. The location of the mask is usually given relative to the node considered. A two-dimensional spatially structured network is shown in Figure 1.2, where a square mask and a Gaussian kernel are used.

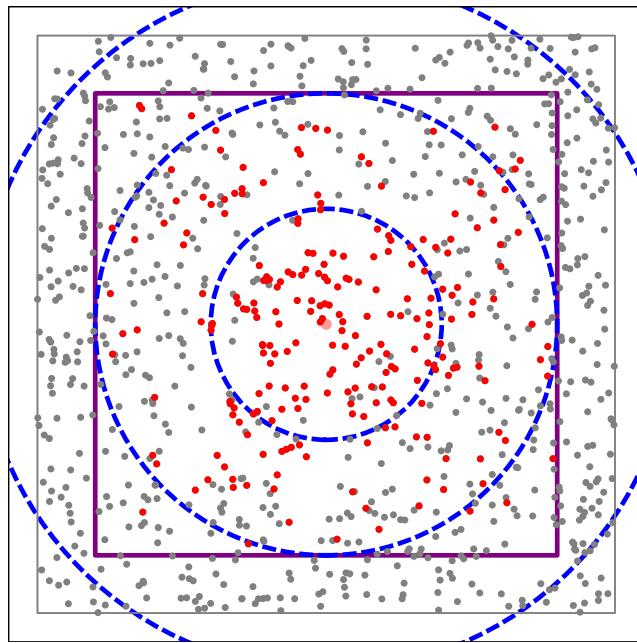
## Aims and organization of this thesis

The aim of this thesis is to develop statistical tests for the probabilistic connection algorithms used in neural network simulators, both for networks with and without spatial structure. Further, an automated test suite that can run relatively quickly, with a low rate of false positives, possibly as part of the automated testing in a continuous integration system, is developed. To demonstrate the utility of the test procedures, they are used to test the connection algorithms of NEST. It is worth emphasizing that the test suites can easily be modified to work with data obtained from any other simulator that shares the basic connection schemes.

Chapter 2 will start with a short introduction to the statistical tests that will be used later, and introduce the concept of a two-level test procedure.



**Figure 1.1:** Example of how a random divergent network might look. All source nodes have an out-degree of 3, while the target nodes have different in-degrees.



**Figure 1.2:** Example of how a spatially structured network might look. 1,000 nodes are scattered across the layer. The centered source node is connected to the red nodes, but not the grey. The mask is shown in purple, and the blue rings mark  $\sigma$ ,  $2\sigma$  and  $3\sigma$  from the Gaussian kernel.

In Chapter 3, a test procedure for random connections with predefined in- or out-degree is developed. NEST-specific implementations are discussed, and results are presented, both for convergent (Section 3.1) and divergent (Section 3.2) connections. In Section 3.3, an automated test procedure is proposed and implemented for NEST.

Networks with spatial structure are discussed in Chapter 4. Test strategies are developed and implemented for NEST, and results are presented, first for two-dimensional space in Section 4.1, then for three dimensions in Section 4.2. A general discussion about findings and perspectives for future research is found in Chapter 5.

As the tests proposed in this thesis are implemented for NEST, there will be references to functions in NEST. The reader might therefore want to take a look at the NEST tutorial found in Appendix A before proceeding.



# Chapter 2

## Statistical testing

The *p*-value approach to statistical hypothesis testing consists of the following five steps ([Ewens and Grant 2004](#)).

**Step 1** is to state the *null hypothesis* ( $H_0$ ) and the *alternative hypothesis* ( $H_1$ ). The aim is to either reject or accept  $H_0$ . Accepting  $H_0$  does not mean that it is necessarily true, only that there was insufficient evidence against it.

**Step 2** is to determine the *level of significance*  $\alpha$ . This is the probability of making a *type I error*, i.e., rejecting  $H_0$  when it is true. We obviously want this probability to be small, but a too small  $\alpha$  will increase the probability  $\beta$  of making a *type II error*, i.e., accepting  $H_0$  when it is false. The choice of  $\alpha$  will therefore depend on the situation and what type of error is most important for us to avoid.

**Step 3** is to decide on a *test statistic*  $T$ . Which test statistic is used depends on the situation.

**Step 4** is to compute the value  $t_{\text{obs}}$  of this test statistic from the observations.

**Step 5** is to compute the *p*-value, i.e., the probability that the test statistic  $T$  would have a value at least as extreme as the observed value  $t_{\text{obs}}$  under  $H_0$ . If the alternative hypothesis corresponds to large values of  $T$ , the *p*-value is calculated as  $\text{Prob}(T \geq t_{\text{obs}})$ . Thus, roughly speaking, a small *p*-value indicates that  $H_0$  is unlikely. The *p*-value is compared with the chosen level of significance  $\alpha$  (typically 0.05 or 0.01). If the *p*-value it is smaller,  $H_0$  is rejected, otherwise,  $H_0$  is accepted.

When the test statistic is continuous, the *p*-value is a continuous, random variable with a uniform distribution  $\mathcal{U}(0, 1)$  under  $H_0$  ([Ewens and Grant 2004](#)),

meaning it will satisfy the equation

$$\text{Prob}(p\text{-value} \leq x | H_0 \text{ true}) = x \quad (2.1)$$

for  $x \in [0, 1]$ . When the test statistic is *discrete*, the  $p$ -value is also discrete, but it is *not* discrete *uniform* under  $H_0$ . The reason is that a  $p$ -value that satisfies the equation above might not exist. Instead, the  $p$ -value will satisfy

$$\text{Prob}(p\text{-value} \leq x | H_0 \text{ true}) \leq x. \quad (2.2)$$

The deviation from the uniform distribution is large when data is sparse, but for larger data sets, the deviation is often negligible.

In certain situations, a *two-level test procedure* is advantageous. First, the expected distribution is compared with the empirical distribution observed, using some goodness-of-fit (GOF) tests. In accordance with classical hypothesis testing, a failure to pass such a test at some level of significance  $\alpha$  would result in the rejection of the null hypothesis  $H_0$  that the empirical frequencies follow the expected distribution, with a probability  $\alpha$  of making a type I error. Successfully passing a test would typically cause  $H_0$  to be accepted, with a probability  $\beta$  of making a type II error.  $\beta$  is often quite large. We might therefore want to run the test several times before we feel confident that  $H_0$  can be accepted. When doing experiments on a computer, the cost of re-running the test is usually small. We can, therefore, run the experiment a large number of times, and check whether the test fails the expected fraction  $\alpha$  of the tests. But this way we would lose a lot of information, and it is arguably not the best approach in this situation. In the case where the  $p$ -value is (sufficiently) uniform under  $H_0$ , we can instead test the observed  $p$ -values generated against the expected  $\mathcal{U}(0, 1)$ . This approach, sometimes referred to as a *two-level test procedure*, is similar to the one used by [L'Ecuyer and Simard \(2007\)](#) to test PRNGs. The test of the  $p$ -values will of course have its own  $p$ -value, which in turn could be compared to the expected uniform distribution using some GOF test, and so on, repeating endlessly. Nevertheless, it should suffice to do one test against uniformity, and compare the resulting  $p$ -value to some predefined level of significance  $\alpha$ . The main advantage of this two-level approach is that if a connection algorithm passes the test, our confidence in it will be much greater than if only a single test was performed.

The GOF test used in the first step of the procedure will vary depending on the data. The tests we will use are introduced below. For the second step (the test of GOF of the  $p$ -values to  $\mathcal{U}(0, 1)$ ) the Kolmogorov-Smirnov test is used.

## 2.1 Pearson's chi-squared test

The following section is based on *A Guide to Chi-Squared Testing* ([Greenwood and Nikulin 1996](#)). Pearson's chi-squared test, sometimes ambiguously referred

to as “the chi-squared test”, is a test of the null hypothesis that there is a good fit between some theoretical distribution and the observed data.

Consider an experiment, or *trial*, resulting in one outcome,  $A_i$ , of  $r$  possible outcomes,  $A_1, A_2, \dots, A_r$ . The probability of outcome  $A_i$  is  $p_i$ . We conduct  $n$  independent trials, indexed  $k = 1, 2, \dots, n$ . We define a random variable

$$\mu_{ki} = \begin{cases} 1 & \text{if outcome } A_i \text{ occurred in } k\text{th trial} \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

for  $i = 1, 2, \dots, r$ . We now define a vector

$$\boldsymbol{\mu}_k = (\mu_{k1}, \mu_{k2}, \dots, \mu_{kr}), \quad k = 1, 2, \dots, n \quad (2.4)$$

which, for each trial  $k$ , describes which outcome occurred. It has only one nonzero component, the one indexed  $i$ , which is equal to 1. The *frequency* of outcome  $A_i$  is defined as

$$\nu_i = \sum_{k=1}^n \mu_{ki}. \quad (2.5)$$

All these frequencies must satisfy the condition  $\nu_1 + \nu_2 + \dots + \nu_r = n$ . The *vector of frequencies* becomes

$$\boldsymbol{\nu} = \sum_{k=1}^n \boldsymbol{\mu}_k = (\nu_1, \nu_2, \dots, \nu_r). \quad (2.6)$$

The expected frequency of a single outcome  $A_i$  is  $E(\nu_i) = np_i$ , with a variance  $\text{var}(\nu_i) = np_i(1 - p_i)$ . This matches that of the binomial distribution, and we might therefore be fooled into thinking that the vector of frequencies  $\boldsymbol{\nu}$  is binomially distributed. This is *not* the case, as the individual frequencies are *dependent* random variables with a restrained sum equal to  $n$ . Instead, the vector of frequencies  $\boldsymbol{\nu}$  has a *multinomial* distribution, with parameters  $n > 0$  and  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ , where  $0 \leq p_i \leq 1$  and  $\sum p_i = 1$ .

The probability mass function (PMF) of the multinomial distribution is given by

$$\text{Prob}(\boldsymbol{\nu} = \mathbf{x}) = \text{Prob}(\nu_1 = x_1, \nu_2 = x_2, \dots, \nu_r = x_r) \quad (2.7)$$

$$= \frac{n!}{x_1! \dots x_r!} p_1^{x_1} \dots p_k^{x_k}, \quad (2.8)$$

where  $\mathbf{x} = (x_1, \dots, x_r)$  is any vector of integers with  $0 \leq x_i \leq n$  and  $\sum_{i=1}^r x_i = n$ . The vector of expected frequencies is

$$E(\boldsymbol{\nu}) = n\mathbf{p} = (np_1, np_2, \dots, np_r). \quad (2.9)$$

In the special case of the multinomial distribution where all the  $p_i = p = \frac{1}{r}$  are equal, the PMF can be written as

$$\text{Prob}(\boldsymbol{\nu} = \mathbf{x}) = \frac{n!}{x_1! \dots x_r!} p^{x_1 + \dots + x_r} = \frac{n!}{x_1! \dots x_r!} p^n, \quad (2.10)$$

and the vector of expected frequencies becomes

$$\mathbb{E}(\boldsymbol{\nu}) = n\mathbf{p} = (np, np, \dots, np). \quad (2.11)$$

To test whether the observed frequencies  $\nu_i$  match the assumed multinomial distribution, we can define the null hypothesis

$$H_0 : p_i = p_i^{(0)}, \quad i = 1, 2, \dots, r. \quad (2.12)$$

We can then use Pearson's chi-squared test. Pearson's test statistic is defined as

$$X^2 = \sum_{i=1}^r \frac{(\nu_i - np_i^{(0)})^2}{np_i^{(0)}} = \sum_{i=1}^r \frac{\nu_i^2 - 2\nu_i np_i^{(0)} + (np_i^{(0)})^2}{np_i^{(0)}} \quad (2.13)$$

$$= \sum_{i=1}^r \frac{\nu_i^2}{np_i^{(0)}} - 2 \sum_{i=1}^r \nu_i + n \sum_{i=1}^r p_i^{(0)} = \sum_{i=1}^r \frac{\nu_i^2}{np_i^{(0)}} - 2n + n \quad (2.14)$$

$$= \frac{1}{n} \sum_{i=1}^r \frac{\nu_i^2}{p_i^{(0)}} - n. \quad (2.15)$$

For the special case where all the  $p_i^{(0)} = p^{(0)} = \frac{1}{r}$  are equal, this simplifies further to

$$X^2 = \frac{r}{n} \sum_{i=1}^r \nu_i^2 - n. \quad (2.16)$$

When the sample size is large, this statistic has an asymptotic chi-squared ( $\chi^2$ ) distribution with  $r - 1$  degrees of freedom. The reduction of 1 degree of freedom is because there is one constraint, namely that the frequencies have to sum to  $n$ . If the expected frequencies  $np_i^{(0)}$  are too small (typically  $< 5$ ) the chi-squared distribution will not be a good approximation to the distribution of the test statistic. The individual trials are assumed to be independent.

Knowing the distribution, the  $p$ -value, defined as the probability of finding a test statistic as large or larger than the observed, i.e.,  $\text{Prob}(\chi^2 \geq X^2)$ , can be calculated. A perfect fit, i.e.,  $\nu_i = np_i^{(0)}$  for all  $i$ , will give a  $X^2 = 0$ , making the  $p$ -value 1. A poor fit will give a large  $X^2$ , and therefore  $p$ -value close to (but never equal to) 0. Thus, the  $p$ -value from a chi-squared test can be seen as a measure of the goodness-of-fit of the data to the expected distribution, a property we will later exploit.

## 2.2 The Kolmogorov-Smirnov test

The following section is based on *Advanced statistics from an elementary point of view* (Panik 2005). Pearson's chi-squared test is well suited for analyzing categorical data or data that naturally fall into distinct groups or bins. To use the chi-squared test on *continuous* data, we would have to group the data into somewhat arbitrarily sized bins. Because the expected frequencies of these bins cannot be too low, there is a lower limit to the bin sizes, and an upper limit to the number of bins. This necessarily causes information to be lost. The chi-squared test is therefore not ideal for continuous data. The Kolmogorov-Smirnov (KS) test is a GOF test that does not require grouping of the sample data, and is therefore much better suited for continuous data.

Let  $X$  be a random variable, and  $x_i$ ,  $i = 1, 2, \dots, n$  be an *ordered* set of  $n$  realizations of  $X$ . We wish to determine if  $X$  has a specific hypothesized distribution, or, in other words, if the observed data  $x_i$  stems from a population with a cumulative distribution function (CDF)  $F(x)$  that equals our hypothesized CDF  $F_0(x)$ . We formulate our null hypothesis,

$$H_0 : F(x) = F_0(x). \quad (2.17)$$

For a *two-sided* KS test, the alternative hypothesis becomes

$$H_1 : F(x) \neq F_0(x). \quad (2.18)$$

Our sample data will have an *empirical distribution function* (EDF)  $S_n(x)$  that converges to the true CDF  $F(x)$  for large  $n$ .  $S_n(x)$  equals the proportion of realizations  $x_i$  that are below  $x$ . It is therefore a discrete function, increasing stepwise by  $1/n$  at each  $x = x_i$ . It can be defined as

$$S_n(x) = \begin{cases} 0, & \text{for } x < x_1 \\ \frac{i}{n}, & \text{for } x_i \leq x < x_{i+1}, i = 1, 2, \dots, n-1 \\ 1, & \text{for } x \geq x_n. \end{cases} \quad (2.19)$$

The Kolmogorov-Smirnov test statistic is defined as the *supremum* of the absolute difference between  $S_n(x)$  and  $F_0(x)$  for all  $x$ ,

$$D_n = \sup_x |S_n(x) - F_0(x)|, \quad (2.20)$$

or simply the greatest distance between the two. If  $H_0$  is true, we expect  $D_n$  to be small. The sampling distribution of  $D_n$  is known, and we can compare our value with this distribution and calculate a *p*-value.

As an example of the usage of the KS test we can test the output of a PRNG, a series of numbers in the half-open interval  $[0, 1)$ , against the expected

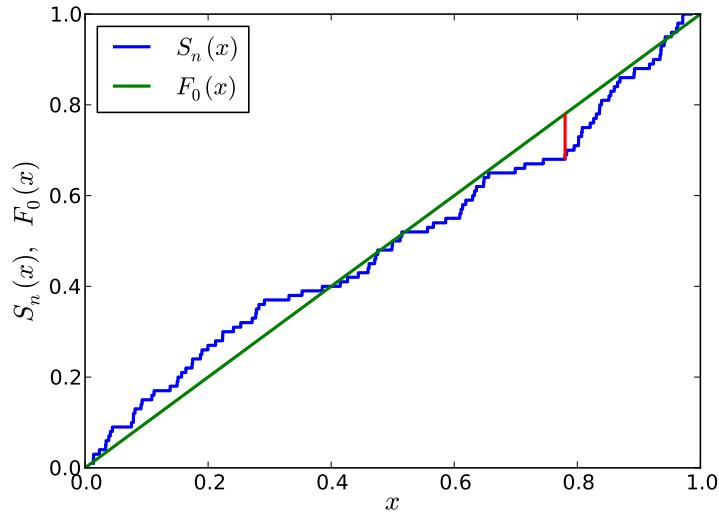
uniform distribution. Then, under  $H_0$ ,  $X \sim \mathcal{U}(0, 1)$ . The expected CDF becomes

$$F_0(x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } 0 \leq x \leq 1 \\ 1, & \text{for } x > 1. \end{cases} \quad (2.21)$$

Having obtained a set of pseudorandom numbers  $x_i$ , the first step is to order these values in increasing order. The EDF can then be found from the definition of  $S_n$  in Equation 2.19, and the test statistic  $D_n$  is found by Equation 2.20. In Figure 2.1 the red line marks the greatest distance  $D_n = 0.11$  between  $S_n(x)$  and  $F_0(x)$  for a set of 100 pseudorandom numbers supposedly drawn from  $\mathcal{U}(0, 1)$ .

The  $p$ -value, i.e., the probability of observing a more extreme value of the KS test statistic for  $n = 100$ , can be shown to be 0.16. There is in other words no evidence to support a rejection of the null hypothesis that the  $x_i$  are drawn from  $\mathcal{U}(0, 1)$  at any meaningful level of significance  $\alpha$ .

Unlike the chi-squared test, where the test statistic only approximates the chi-squared distribution, the KS test is an exact test, meaning it can be used for small  $n$  as well as large.



**Figure 2.1:** Empirical distribution function  $S_n(x)$  (blue line) of  $n = 100$  pseudorandom numbers from PRNG, supposedly drawn from a uniform distribution, and the cumulative distribution function  $F_0(x)$  (green line) expected under  $H_0$ . The red line marks the KS test statistic  $D_n$ , the greatest distance between the two lines.

## 2.3 The $Z$ -test

The following section is based on *Advanced statistics from an elementary point of view* (Panik 2005).

The  $Z$ -test is a test of the hypothesis that a sample is drawn from a normal population with the mean  $\mu_0$ . Let  $(X_1, X_2, \dots, X_n)$  be a set of  $n$  random variables, drawn from a normal population with a known variance  $\sigma^2$ , but an unknown population mean  $\mu$ . The sample mean is denoted  $\bar{X}$ , and the standard deviation of the mean is related to the standard deviation of the population by  $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ .

We wish to determine whether the population mean  $\mu$  equals a hypothesized mean  $\mu_0$ , i.e.,

$$H_0 : \mu = \mu_0. \quad (2.22)$$

We consider here only the two-sided alternative hypothesis,

$$H_1 : \mu \neq \mu_0. \quad (2.23)$$

The test statistic used is the standard score,

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} \quad (2.24)$$

which, under  $H_0$ , has a standard normal distribution  $\mathcal{N}(0, 1)$ . The  $p$ -value of the two-sided  $Z$ -test is the probability of finding a value of  $Z$  as extreme or more extreme than the observed value  $z$  under  $H_0$ , i.e.,

$$p\text{-value} = \text{Prob}(Z \leq -|z|) + \text{Prob}(Z \geq |z|) \quad (2.25)$$

$$= 2\text{Prob}(Z \geq |z|). \quad (2.26)$$

If the data set consists of a single random variable  $X$ , the test statistic becomes

$$Z = \frac{X - \mu_0}{\sigma}. \quad (2.27)$$

The  $Z$ -test can only be used for data that can be approximated by a normal distribution. The central limit theorem is typically invoked to justify the approximation.



# Chapter 3

## Distribution of connections

We will now describe a procedure for testing random convergent and divergent connection algorithms for networks without spatial structure. These connection algorithms were described in the introduction, and an example of a resulting network was shown in Figure 1.1.

### 3.1 Random convergent connections

For each target node `RandomConvergentConnect` iterates over, it randomly draws  $C$  source nodes from the  $r = N_s$  available source nodes and connects to them. Multapses and autapses are allowed, i.e., nodes are drawn with replacement. We are interested in checking whether all the source nodes are drawn with equal probability. In other words, we want to test the observed distribution of connections against the expected uniform distribution. Each drawing of a source node can be thought of as a trial as described in Section 2.1, with exactly *one* outcome  $A_i$ . The total number of trials is  $n = N_t \times C$ , where  $N_t$  is the number of target neurons. After all  $n$  trials, each source node  $i$  will have some out-degree (number of outgoing connections), described by the *frequency of outcome*  $A_i$  as defined in Equation 2.5:

$$\nu_i = \sum_{k=1}^n \mu_{ki}. \quad (3.1)$$

The vector of frequencies,

$$\boldsymbol{\nu} = \sum_{k=1}^n \boldsymbol{\mu}_k = (\nu_1, \nu_2, \dots, \nu_r), \quad (3.2)$$

now contains all observed out-degrees.

In this particular case, the *expected* frequencies are all the same,  $E(\nu_i) = np^{(0)}$ , so the vector of expected frequencies contains  $r$  equal entries,  $E(\boldsymbol{\nu}) = n\boldsymbol{p}^{(0)} = (np^{(0)}, np^{(0)}, \dots, np^{(0)})$ .

To test the hypothesis

$$H_0 : p_i = p_i^{(0)} = p^{(0)} = \frac{1}{r}, \quad i = 1, 2, \dots, r, \quad (3.3)$$

i.e., that source nodes are drawn with equal probability, we use Pearson's chi-squared ( $\chi^2$ ) test. As explained in Section 2.1, the statistic becomes

$$X^2 = \frac{r}{n} \sum_{i=1}^r v_i^2 - n. \quad (3.4)$$

This statistic has an asymptotic chi-squared ( $\chi^2$ ) distribution with  $r-1$  degrees of freedom. Knowing the distribution, the  $p$ -value can be calculated.

Instead of simply rejecting or accepting  $H_0$  based on one  $p$ -value, it is, as argued earlier, better to apply a two-level test, i.e., test multiple  $p$ -values for a uniform distribution. This assumes the  $p$ -value is uniformly distributed under  $H_0$ , which is not strictly true for  $p$ -values coming from a chi-squared test, as  $X^2$  is discrete. To examine the discreteness of  $X^2$  closer, let us change  $\nu_1$  to  $\nu_1 + 1$ , and  $\nu_2$  to  $\nu_2 - 1$ . The resulting change in  $X^2$  is

$$\begin{aligned} \Delta X^2 &= \left[ \frac{r}{n} \left( [\nu_1 + 1]^2 + [\nu_2 - 1]^2 + \sum_{i=3}^r \nu_i^2 \right) - n \right] - \left[ \frac{r}{n} \sum_{i=1}^r \nu_i^2 - n \right] \\ &= \frac{r}{n} ([\nu_1 + 1]^2 + [\nu_2 - 1]^2 - \nu_1^2 - \nu_2^2) \\ &= \frac{2r}{n} (\nu_2 - \nu_1 - 1) \end{aligned}$$

Thus, the smallest non-zero difference between two possible values of  $X^2$  is  $2r/n = 2N_s/(N_t C)$ . As long as  $n = N_t C$  is large enough compared to  $r = N_s$ , therefore, these "jumps" in  $X^2$  are small, and we may treat it as continuous; the effects of small  $N_t C$  is investigated in Section 3.1.2. This means that the two-sided Kolmogorov-Smirnov (KS) test can be used to test the uniformity of the  $p$ -values. The KS test produces a  $p$ -value which, if it is smaller than a chosen significance level  $\alpha$ , leads us to reject  $H_0$ . An advantage of this approach is that, even though Pearson's chi-squared test is one-tailed, a "too good" fit (connections are more evenly distributed than is likely to happen by chance) will be detected, as there will be an excess of large  $p$ -values from the chi-squared tests.

### 3.1.1 Implementation

The test procedure outlined above is implemented as a Python module, included in Appendix B. The module defines a class, `RCC_tester`. The class has two methods, `chi_squared_test` and `two_level_test`, for testing the connections created by `RandomConvergentConnect`.

`chi_squared_test` creates two sets of nodes, `source_nodes` and `target_nodes`, and connects them using `RandomConvergentConnect`. It then runs a chi-squared test on the out-degrees of the source nodes, and returns the test statistic and the  $p$ -value. If the expected frequencies  $np_i$  are too small, results may be unreliable. Thus, if they are smaller than  $e_{\min}$  (10 by default), a warning is displayed.  $e_{\min}$  can be changed. The chi-squared test is implemented using the `chisquare` function from the `scipy.stats` library.

The method `two_level_test` runs `chi_squared_test`  $n_{\text{runs}}$  times, and checks the returned  $p$ -values for uniformity using the two-sided KS test. The resulting KS test statistic and  $p$ -value is returned. The KS test is implemented using the `kstest` function from the `scipy.stats` library. In the main section at the end of the module an example of how to use the `RCC_tester` class is provided.

Between each run of `RandomConvergentConnect`, the network is deleted by calling the function `ResetKernel`, to avoid memory bloat. This also resets the PRNGs. Thus, for each run of `RandomConvergentConnect`, NEST must be given a new set of PRNG seed values. NEST requires one seed value for the global PRNG and one for each per-process PRNG, totaling  $1 + n_{\text{vp}}$  seed values, where  $n_{\text{vp}}$  is the number of virtual processes (VPs) used by NEST. The `chi_squared_test` method takes one argument, a “master seed” `msd`, and the  $1 + n_{\text{vp}}$  PRNGs are seeded with the values  $(\text{msd}, \text{msd}+1, \dots, \text{msd}+n_{\text{vp}})$ . For independent results, `chi_squared_test` should be given a new master seed for each run, differing by at least  $n_{\text{vp}} + 1$ . The `two_level_test` method handles this automatically when running `chi_squared_test`. The first of the master seeds can be passed as an argument `start_seed`. When running `two_level_test` multiple times, `start_seed` should differ by at least  $n_{\text{runs}}(n_{\text{vp}} + 1)$ .

### 3.1.2 Results

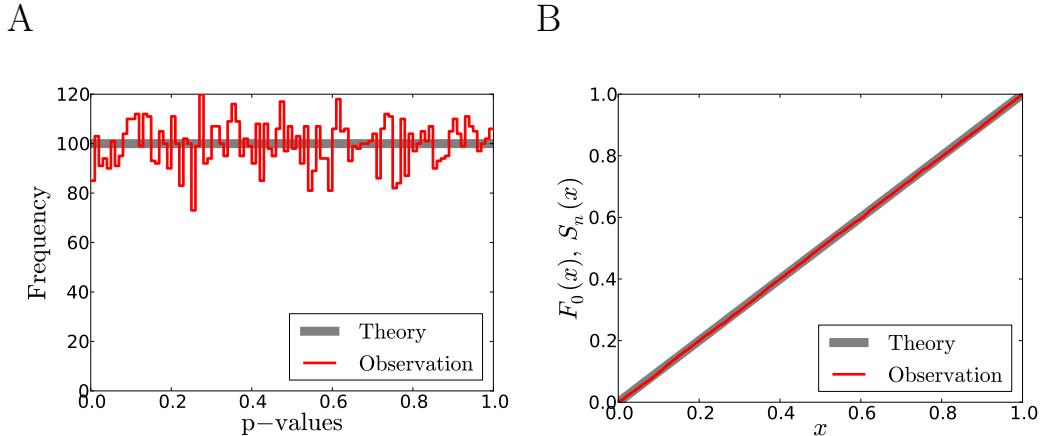
Running the script in Appendix B with the parameter set shown in Table 3.1, we obtain a set of chi-squared  $p$ -values with the distribution shown in Figure 3.1. The KS test of the uniformity of these  $p$ -values results in the KS

Parameter	$N_s$	$N_t$	$C$	$n_{\text{runs}}$	<code>start_seed</code>
Value	1,000	1,000	1,000	10,000	0

**Table 3.1:** Default parameter set for reported results.

test statistic  $D_n = 5.67 \times 10^{-3}$  and a  $p$ -value of 0.905, leading us to accept the null hypothesis  $H_0$  (defined in Equation 3.3) that the nodes were selected with equal probability.

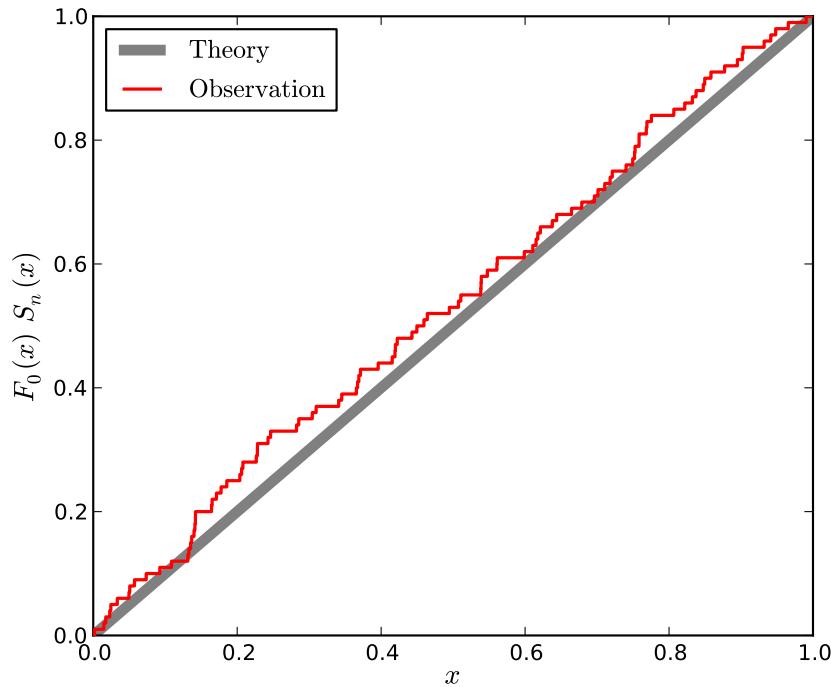
To assess the two-level testing procedure itself, it might be useful to be able to run it on data that we can safely assume to fulfill  $H_0$ . The test



**Figure 3.1:** **A:** Histogram of  $p$ -values from chi-squared goodness-of-fit tests of  $n = 10,000$  individual networks, combined into 100 bins (red) together with the expected uniform distribution (grey). **B:** The corresponding empirical distribution  $S_n(x)$  (red) and expected cumulative distribution  $F_0(x)$  (grey).

script allows us to do this with the optional argument `control=True` passed to `two_level_test`. NEST’s connection algorithm is then swapped with a simple “control algorithm” that creates data that matches the multinomially distributed vector of degrees we would expect to get from NEST. It runs faster than the actual connection algorithm, and can therefore generate a larger data set in the same period of time. As mentioned briefly in Section 3.1, the  $p$ -value from a chi-squared test is not truly a continuous variable. This distinctness might cause the two-level test to give a left-skewed distribution of  $p$ -values for certain combinations of parameters. The control algorithm can be used to investigate this. Running the two-level test procedure on the control algorithm, with  $n_{\text{runs}} = 1,000$ , repeated 100 times, each time with a different starting seed, yields the EDF of  $p$ -values in Figure 3.2. These  $p$ -values appear to be uniformly distributed (a KS test of uniformity results in the  $p$ -value 0.468). This is an important result, as the uniformity of the  $p$ -values from the two-level test procedure under  $H_0$  is a prerequisite for drawing conclusions based on them.

The jumps of the chi-squared  $p$ -value will be large when the expected out-degree  $N_t C / N_s$  is small, i.e., for small  $N_t$  and  $C$ , and large  $N_s$ . The effect of small degrees is investigated. Figure 3.3 shows the EDF of  $p$ -values with four different combinations of small values for these parameters. Jumps in the  $p$ -values are clearly seen. As expected, the jumps grow larger with a smaller expected out-degree. The  $p$ -values from the two-level test corresponding to the EDFs in Figure 3.3A, 3.3B, and 3.3C, are  $2.47 \times 10^{-25}$ ,  $6.47 \times 10^{-5}$ , and 0.0162, respectively. These  $p$ -values are clearly left-skewed, even though the data tested does fulfill  $H_0$ . In the last figure, 3.3D, the expected degree is 100,



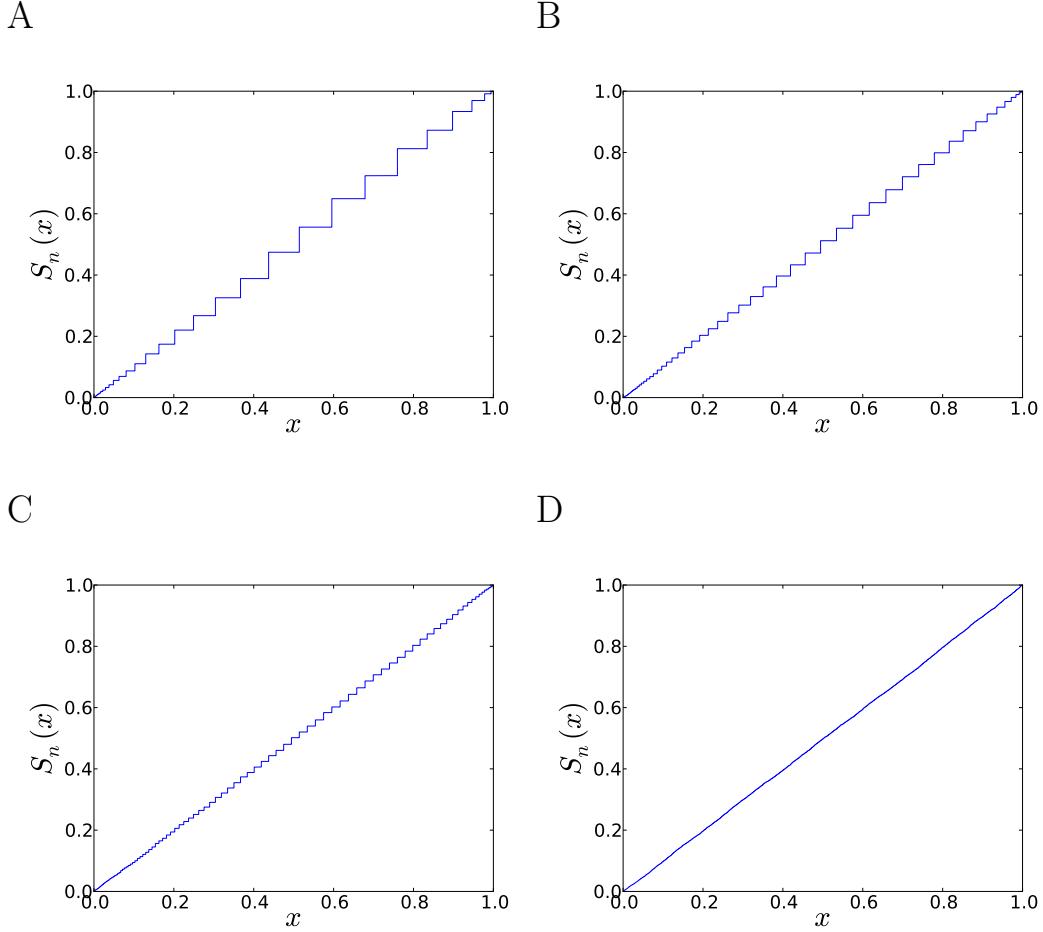
**Figure 3.2:** EDF of 100  $p$ -values from two-level test procedure. For each run, the two-level test procedure generates  $n = 1,000$  networks using the control algorithm, runs a chi-squared GOF tests on the vector of degrees, and tests the resulting  $p$ -values for uniformity using the KS test.

and the  $p$ -value resulting from the two-level test is 0.237. Re-running multiple times reveals that the fraction of  $p$ -values below  $\alpha$  is close to  $\alpha$ . Thus, an expected degree of about 100 seems to suffice to produce  $p$ -values that can be used to draw meaningful conclusions.

Note that even though the two-level test will result in left-skewed  $p$ -values for small expected degrees, the fraction of  $p$ -values from a chi-squared test that lie below some level of significance  $\alpha$  will be very close to  $\alpha$ , partly because the biggest jumps of the EDF are in the middle region, while in the lower part, close to 0, the curves are fairly smooth (see Figure 3.3). In other words, a two-level test might not be reliable for small expected degrees ( $\lesssim 100$ ), but a “one-level” chi-squared test can be used with much smaller expected degrees ( $\sim 5$ ).

### Sensitivity

To assess the sensitivity of the two-level test procedure, i.e., its ability to detect small errors and biases, various errors and biases were deliberately introduced into the data. Some of these are described below, as well as the resulting  $p$ -values. We emphasize that these  $p$ -values are only examples. Rerunning the



**Figure 3.3:** EDFs of 10,000  $p$ -values from chi-squared tests with four different combinations of values for the parameters  $N_s$ ,  $N_t$ , and  $C$ . In **A**,  $N_s = 10$ ,  $N_t = 5$ , and  $C = 5$ , resulting in an expected out-degree of 2.5. The  $p$ -values are clearly distinct, with fairly large jumps between adjacent values, especially for medium to large values. In **B**,  $N_s = 10$ ,  $N_t = 10$ , and  $C = 5$ . This gives an expected degree of 5, resulting in smaller jumps. In **C**,  $N_s = N_t = C = 10$ , resulting in an expected degree of 10, and the jumps are smaller yet. In **D**,  $N_s = 10$ ,  $N_t = 100$ , and  $C = 10$ , giving an expected degree of 100. The jumps are no longer visible.

tests with the same biased algorithm and the same parameters, but with a different PRNG seed value, will result in a different  $p$ -value, possibly one that differs by quite a bit. Thus, they are only meant to give an indication of how well the bias is detected. Unless otherwise stated, the parameter values listed in Table 3.2 are used.

The first bias that was deliberately introduced was a simple right-skewing of the data. The degree of the first half of the nodes was reduced by one, and the degrees of the second half was increased by one. This was easily detected by the test procedure ( $p = 1.71 \times 10^{-5}$ ). In Figure 3.4, the distribution of 1,000

Parameter	$N_s$	$N_t$	$C$	$n_{\text{runs}}$	<code>start_seed</code>	<code>control</code>
Value	1,000	1,000	100	1,000	0	True

**Table 3.2:** Default parameter set for sensitivity testing.

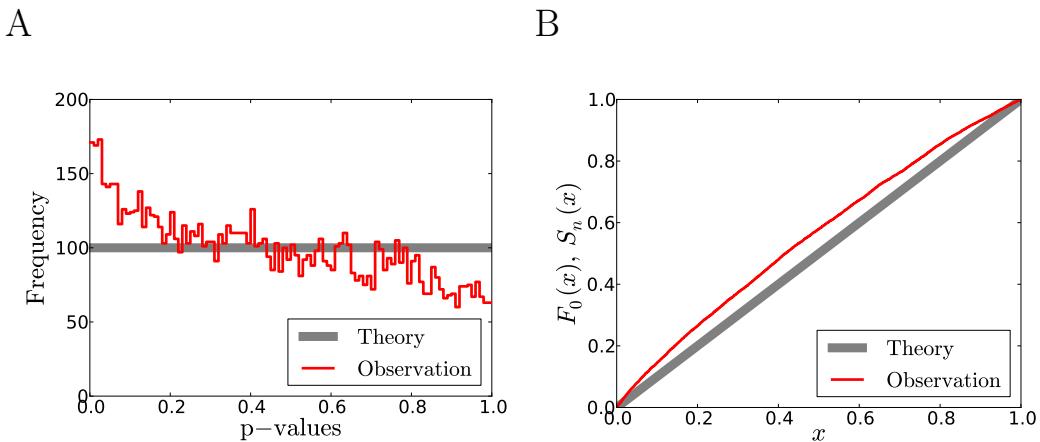
$p$ -values are shown. It is clear from the figure that a single chi-squared test would not in any consistent way detect the bias. Only when we accumulate a large number of  $p$ -values is the trend obvious.

When  $C$  was increased to 1,000, the bias was not detected ( $p = 0.255$ ). A large  $C$  is clearly not always advantageous for detecting small biases, especially biases that do not increase with  $C$ . Only after  $n_{\text{runs}}$  was increased to 10,000 was the bias again detected ( $p = 0.00428$ ).

A second bias was introduced by decreasing the degree of every source node with an even-numbered index by one, and increasing the degree of every source node with an odd-numbered index by one. This was also easily detected ( $p = 7.13 \times 10^{-6}$ ).

A third bias was caused by increasing by one the degrees below the 5th percentile, and decreasing by one the degrees above the 95th percentile, thereby making the vector of frequencies a slightly “too good” fit to the theoretical distribution. This change was detected, both with  $C = 100$  ( $p = 8.02 \times 10^{-14}$ ) and  $C = 1,000$  ( $p = 0.00231$ ), for  $n_{\text{runs}} = 100$ , as well as for larger  $n_{\text{runs}}$ .

A fourth bias was introduced by moving all the connections from one of the  $N_s$  nodes to another node. This kind of error could easily be introduced by confusion between the zero-based numbering typically used in computer science and the one-based numbering used in everyday circumstances. The

**Figure 3.4:** Histogram (A) and CDF (B) showing the distribution of  $p$ -values from chi-squared tests after introducing an error into the connection algorithm. Expectation is shown in grey.

error was easily detected (the  $p$ -value was reported as 0.0 as it was too small for Python’s built-in floating point data type to handle). Increasing  $N_s$  to 1,000, the error was still detected at  $\alpha = 0.05$  ( $p = 0.0366$ ).

A fifth, very small bias was introduced by increasing the degree of one of the source nodes by one, and decreasing the degree of another node by one. The bias was not detected at  $\alpha = 0.05$  with  $C = 100$  and  $n_{\text{runs}} = 10,000$  ( $p = 0.0740$ ), nor with  $C$  reduced to 10. Reducing  $N_s$  and  $N_t$  to 100, the bias was detected at  $\alpha = 0.05$  ( $p = 0.0189$ ). A smaller network clearly increases the sensitivity to certain types of biases, especially biases that do not increase with network size.

Several similar tests were run. Generally the two-level test procedure seems quite sensitive to small biases. The sensitivity obviously increases with  $n_{\text{runs}}$ , but not necessarily with  $C$ ,  $N_t$  and  $N_s$ . The effect of these parameters on the sensitivity depends on the nature of the bias we wish to detect. It might therefore be a good idea to run the procedure with different sets of parameters.

## 3.2 Random divergent connections

`RandomDivergentConnect` works in very much the same way as `RandomConvergentConnect`, except that the source nodes are now iterated over, and target nodes are randomly drawn and connected, meaning that the out-degrees of the source nodes are now all  $C$ , while the in-degrees of the target nodes will vary. It might therefore seem superfluous to test both functions. Running NEST with multiple VPs, however, the exact implementation of the connection algorithm is not the same for the two functions. This is because different nodes are handled by different VPs, and information about connections is handled by the same VP by which the *target* node is handled (Plessner, Eppler, Morrison, Diesmann, and Gewaltig 2007). To minimize the amount of inter-VP communication necessary, each VP is given its own PRNG. `RandomDivergentConnect` uses the global PRNG when drawing connections, while `RandomConvergentConnect` uses the per-VP PRNGs. The result is different connectivity patterns, even with the same master seed value. Both function should therefore be tested with multiple VPs.

### 3.2.1 Implementation

The Python module for testing `RandomDivergentConnect`, found in Appendix C, is very similar to the module for testing `RandomConvergentConnect`. A class `RDC_tester`, with the two methods `chi_squared_test` and `two_level_test`, is defined. These methods take the same parameters as their counterparts in the convergent case. As before, the usage is demonstrated in the main section of the module.

### 3.2.2 Results

Running the script in Appendix C with the same parameter set used for random convergent connections, listed in Table 3.1, results in a  $p$ -value of 0.905. This matches exactly the  $p$ -value we got in Section 3.1.2 after running the test script in Appendix B with the same set of parameters. In fact, the distribution of connections among source nodes for a network created by `RandomConvergentConnect` will exactly match the distribution of connections among target nodes for a network created by `RandomDivergentConnect`, when NEST is run with one VP. For multiple VPs, however, the resulting distribution of connections is not the same. The test class in the test script can be instantiated with the additional argument `threads`, causing NEST to operate with the specified number of local threads. With NEST running as a single process, the number of VPs will equal the number of local threads. Running the script in Appendix C with the same parameter set as before (listed in Table 3.1), but with the extra argument `threads = 2`, we obtained a  $p$ -value of 0.931, consistent with the expected multinomial distribution of in-degrees.

## 3.3 Automated test procedure

We will now describe how to turn a variant of the test procedure described in the previous sections into an automated test, implemented as a unit test. This places some practical limitations on the test procedure. The test must not take too long to run. It should also not have a too high rate of false positives (type I error). At the same time we do of course not want to lose too much sensitivity.

A quick and efficient approach would be to generate one single network and run a chi-squared test on the distribution of connections with some relatively low level of significance  $\alpha$ , say, 0.01, as described earlier, but even with this low level of significance, we would still see false positives for about 1% of the test runs. Reducing  $\alpha$  further would result in a lower sensitivity. On the other hand, running the more thorough but slow two-level test procedure described in Section 3.1 every time is quite time-consuming. Instead, an adaptive approach is proposed here, similar to the one used by L'Ecuyer and Simard (2007). First, a single network is generated, and a chi-squared test is performed on the distribution of connections. If the resulting  $p$ -value is deemed too extreme, a larger number  $n_{\text{runs}}$  of networks is generated, and a more thorough two-level test is performed, either confirming or allaying our suspicion. This will allow us to require a fairly high level of significance for the chi-squared test, thereby keeping the sensitivity high, while at the same time having a low rate of false positives.

As was shown in Section 3.1.2, not all errors and biases are best detected

with a large network or with large degrees, as this might hide certain types of small biases. Tests should therefore be run with different values of  $N_s$ ,  $N_t$ , and  $C$ . For a single chi-square test, the network can be quite small, and the only limit is that the expected degree must not be below  $e_{\min}$ , but the two-level test procedure is more sensitive, and the expected degree should not be much lower than 100. Both `RandomConvergentConnect` and `RandomDivergentConnect` should be tested. Running tests both with a single VP and with multiple VPs might also be a good idea, as the implementation of the connection algorithms are somewhat different for multiple VPs (see Section 3.2).

For the  $p$ -values from the chi-squared test, both very low and very high values are considered suspicious. Let  $\alpha_{1,\text{lower}}$  and  $\alpha_{1,\text{upper}}$  be the values below and above which, respectively,  $p$ -values are deemed suspicious. Under  $H_0$ , suspicious values will occur for a fraction  $\alpha_1 = \alpha_{1,\text{lower}} + (1 - \alpha_{1,\text{upper}})$  of the tests performed. Whenever such an extreme value is encountered, the two-level test is performed. Let  $\alpha_2$  be the value below which the  $p$ -value from the KS test is considered too extreme, and  $H_0$  is rejected. The total fraction of false positives will then be

$$\alpha = \alpha_1 \times \alpha_2 = (\alpha_{1,\text{lower}} + 1 - \alpha_{1,\text{upper}}) \times \alpha_2 \quad (3.5)$$

With  $\alpha_{1,\text{lower}} = 0.025$ ,  $\alpha_{1,\text{upper}} = 0.975$ , and  $\alpha_2 = 0.05$ ,  $\alpha$  becomes 0.0025. The choice of these critical values, as well as the test parameters ( $N_s$ ,  $N_t$ ,  $C$ ,  $n_{\text{runs}}$ ) will depend on the intended usage of the unit test, the computing power of the system it will run on, the maximum time the test can be allowed to take, the desired fraction of false positives, etc.

### 3.3.1 Implementation

An implementation of the automated test procedure can be found in Appendix D. It is implemented using the Python unit testing framework `unittest`. The test scripts for `RandomConvergentConnect` and `RandomDivergentConnect`, found in Appendix B and C, are imported to avoid code duplication.

The test is repeated three times for each of `RandomConvergentConnect` and `RandomDivergentConnect`, once with a small network, once with a larger network, and once with multiple VPs, giving a total of six test cases.  $n_{\text{runs}}$  can have different values for each test, as we might want to run a test on a small network a larger number of times than a test on a large network. The critical values are  $\alpha_{1,\text{lower}} = 0.025$ ,  $\alpha_{1,\text{upper}} = 0.975$ , and  $\alpha_2 = 0.05$ . We thus expect a fraction 0.0025 of the tests to fail. When running the test suite with the six test cases 500 times, giving a total of 3,000 test runs, 10 failures were encountered. This is close to the  $\sim 0.0025 \times 3000 = 7.5$  false positives we would expect.

When the six one-level chi-squared tests are passed, the suite takes less than a minute to complete on most systems. In the event that all six one-level

tests fail (which will happen with a probability of  $0.05^6 = 1.5625 \times 10^{-8}$  under  $H_0$ ), and the two-level test is invoked, the entire suite might take a few minutes to complete, depending on hardware, installed version of NEST etc.



# Chapter 4

## Spatially structured networks

We will now describe the procedure for testing the probabilistic connection algorithms for spatially structured networks. We briefly discussed spatially structured networks in the introduction, and an example was shown in Figure 1.2. In NEST, these connection patterns can be created between two- or three-dimensional layers using the function `ConnectLayers` from the `Topology` module ([Plessner and Enger 2012](#)).

### 4.1 Two-dimensional space

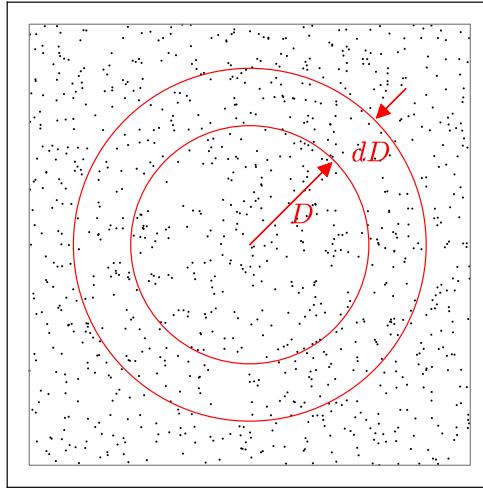
Concepts and derivations in this section are based on [Kriener \(2012\)](#). Let one node be centered on a quadratic  $L \times L$  layer. It can connect to  $N$  other nodes, uniformly distributed on the layer, as long as they are inside the mask. The mask has the same size as the layer, and for now, the same position. The probability of making a connection is given by a distance-dependent connection probability (kernel). Let  $(x_i)_m$  be the  $m$ th component of the coordinate vector for node  $i$ . Let

$$D_{ij} = \sqrt{\sum_{m=1}^k (\Delta x_{ij})_m^2}, \quad (4.1)$$

where  $k = 2$  is the number of dimensions, be the distance between nodes  $i$  and  $j$ . With periodic boundary conditions,

$$(\Delta x_{ij})_m = \begin{cases} |(x_i)_m - (x_j)_m| & \text{for } |(x_i)_m - (x_j)_m| \leq L/2 \\ L - |(x_i)_m - (x_j)_m| & \text{for } |(x_i)_m - (x_j)_m| > L/2 \end{cases} \quad (4.2)$$

We wish to compare the observed distribution of distances between connected nodes with the expected distribution. The expected distribution depends on both on the number of potential targets at a given distance, and the probability of connecting to nodes at that distance (given by the kernel). We



**Figure 4.1:** A ring of radius  $D$  and thickness  $dD$  is placed on a layer with a node density  $\rho_0 = N/L^2$ . The expected number of nodes inside the ring is then  $dN = \rho_0\pi [(D + dD)^2 - D^2]$ .

start by deriving an expression for the number of potential targets at a given distance.

Let  $\rho_0 = N/L^2$  be the average node density on the layer. Now consider a ring like the one in Figure 4.1, with radius  $D$  and thickness  $dD$ . The expected number of nodes in the ring will be

$$dN = \rho_0\pi [(D + dD)^2 - D^2] \quad (4.3)$$

and the density of nodes in the ring is given by  $dN/dD$ . We now define the radial distribution function (RDF) as the limit of the node density for an infinitesimally thin ring

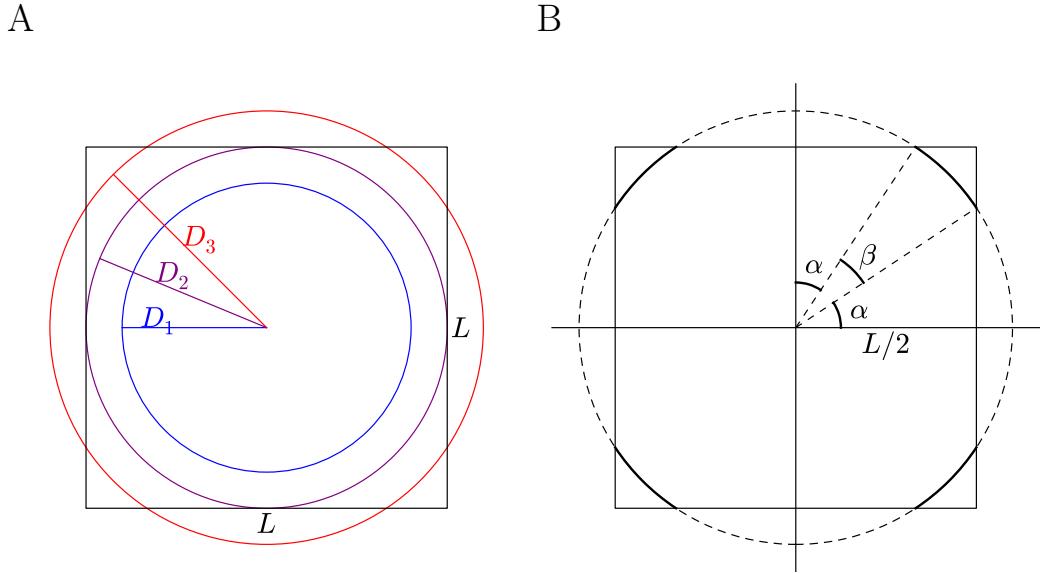
$$\rho(D) = \lim_{dD \rightarrow 0} \frac{dN}{dD} = 2\pi\rho_0 D. \quad (4.4)$$

Thus, for distances  $D \in [0, L/2]$  from the center node, the RDF is proportional to the circumference of a circle with radius  $D$ . For  $D \in (L/2, L/\sqrt{2}]$ , the mask comes into play, as part of the circle of radius  $D$  is outside the mask. This is illustrated in Figure 4.2A. As seen in Figure 4.2B, only a fraction  $2\beta/\pi$  lies inside the mask, where

$$\beta = \frac{\pi}{2} - 2\alpha = \frac{\pi}{2} - 2 \arccos\left(\frac{L}{2D}\right). \quad (4.5)$$

Substituting for  $\beta$  from 4.5 the fraction becomes

$$\frac{2\beta}{\pi} = \frac{\pi - 4 \arccos\left(\frac{L}{2D}\right)}{\pi}. \quad (4.6)$$



**Figure 4.2:** **A:** Illustration of the effect of the mask on the radial distribution function (RDF). At distance  $D_1$ , the RDF is  $\rho_0 2\pi D$ . At distances  $D \geq D_2$ , the mask comes into play. At distance  $D_3$ , for example, part of the circle is outside the mask, and the number of nodes eligible for connecting to thus reduced. **B:** The fraction of a circle with radius  $D$  which is inside the mask is  $2\beta/\pi$ . Adapted from Kriener (2012).

The RDF can therefore be summarized as

$$\rho(D) = \begin{cases} \rho_0 2\pi D & \text{for } 0 \leq D \leq \frac{L}{2} \\ \rho_0 2D \left( \pi - 4 \arccos \left( \frac{L}{2D} \right) \right) & \text{for } \frac{L}{2} < D \leq \frac{L}{\sqrt{2}} \\ 0 & \text{otherwise.} \end{cases} \quad (4.7)$$

Let  $\mathcal{P}(D)$  be a distance-dependent connection probability function (kernel). The probability density function (PDF)  $f(D)$  of distances from the centered node and connected nodes is the normalized product of the RDF and the kernel, i.e.,

$$f(D) = \frac{\rho(D)\mathcal{P}(D)}{\int_0^{L/\sqrt{2}} \rho(R)\mathcal{P}(R) dR}, \quad (4.8)$$

where the denominator is a normalizing constant<sup>1</sup>. The cumulative distribution function (CDF)  $F(D)$  can be obtained by integrating  $f(D)$ , i.e.,

$$F(D) = \frac{\int_0^D \rho(R)\mathcal{P}(R) dR}{\int_0^{L/\sqrt{2}} \rho(R)\mathcal{P}(R) dR}. \quad (4.9)$$

<sup>1</sup> $R$  is used instead of  $D$  in integrals as  $D$  sometimes plays the role of the upper integration limit, for instance in Equation 4.9.

As an example, let  $L = 1$  and the kernel be a linear function of  $D$ ,  $\mathcal{P}(D) = (c - aD)H(c/a - D)$ , where  $H$  is the Heaviside step function. For simplicity we assume that  $c/a < L/2$  so that there are no boundary effects. The PDF then becomes

$$f(D) = \frac{\rho_0 2\pi D(c - aD)}{\int_0^{c/a} \rho_0 2\pi R(c - aR) dR} = \frac{6a^2 D(c - aD)}{c^3}, \quad (4.10)$$

and the CDF becomes

$$F(D) = \frac{6a^2}{c^3} \int_0^D R(c - aR) dR = \frac{a^2 D^2 (3c - 2aD)}{c^3}. \quad (4.11)$$

Using numerical integration, CDFs can be similarly found for other kernels. In Figure 4.3, an exemplary PDF and the corresponding CDF are shown, as well as the connectivity pattern of the network. A Gaussian kernel is used.

Using the two-sided KS test, the observed EDFs can now be compared with these theoretical CDFs, with the null hypothesis  $H_0$  that distances are drawn from the theoretical PDFs. The KS test is the natural choice here because we are looking at the distribution of a continuous parameter, the distance.

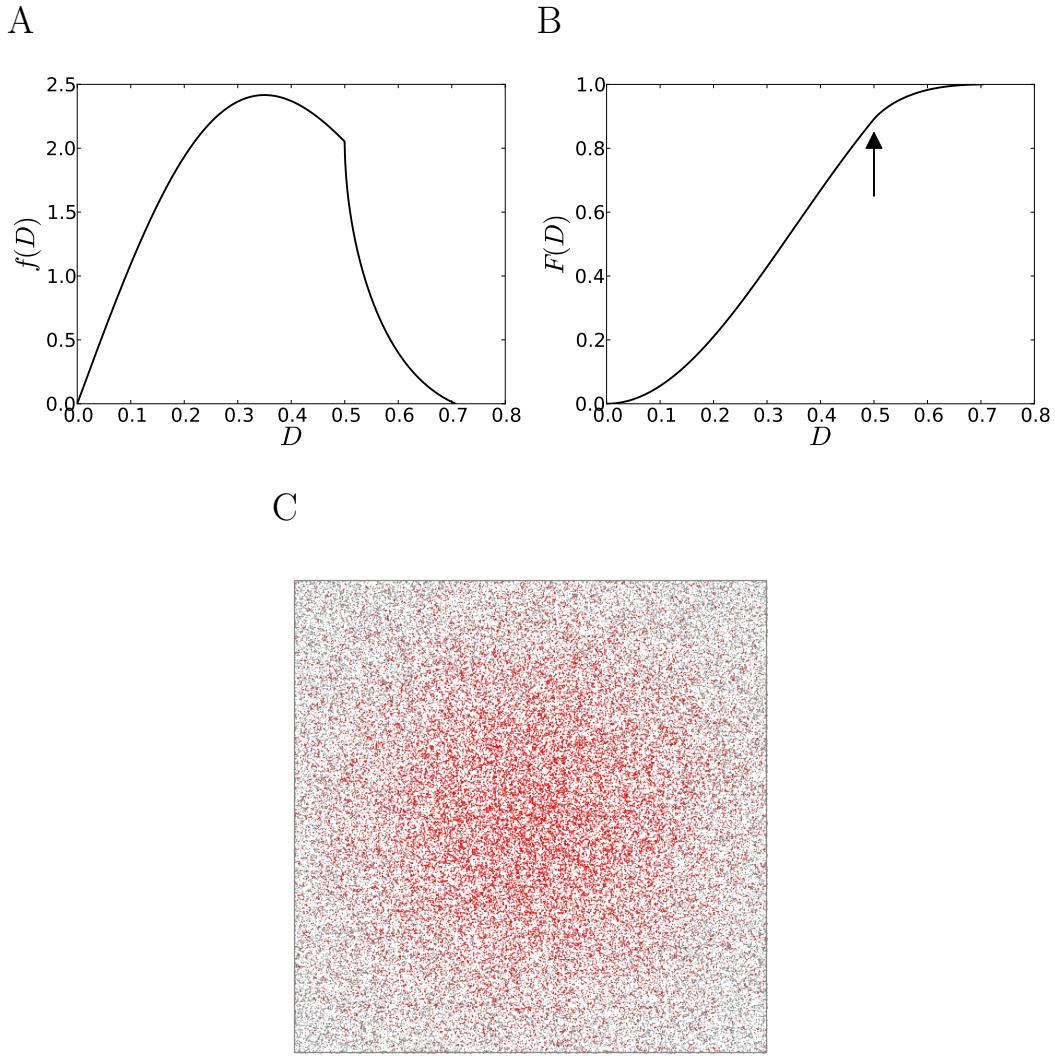
As will be discussed in Section 4.1.2, a few types of errors will not be detected by the KS test of the distribution of source-target distances. A second test is therefore implemented. It simply compares the total number of connections  $C$  with the expected number of connections  $C_0$ .  $C$  will be the sum of  $N$  independent Bernoulli random variables with different success probabilities  $p$  given by the kernel. Its distribution is the Poisson binomial distribution, a generalization of the binomial distribution that does not require all success probabilities to be the same (Wang 1993). For large  $N$  it approximates the normal distribution  $\mathcal{N}(\mu, \sigma^2)$ , with  $\mu = \sum_{i=1}^N p_i$  and  $\sigma^2 = \sum_{i=1}^N p_i(1 - p_i)$ . The Z-test can thus be used as described in Section 2.3. Using the test statistic

$$Z = \frac{C - C_0}{\sigma} = \frac{C - \sum_{i=1}^N p_i}{\sqrt{\sum_{i=1}^N p_i(1 - p_i)}}, \quad (4.12)$$

the two-sided  $p$ -value  $2\text{Prob}(Z \geq |z|)$  can be calculated.

### 4.1.1 Implementation

An implementation of the test procedure described above is found in Appendix E. The main block at the bottom demonstrates the usage. The class `ConnectLayers2D_tester` is first instantiated with the required arguments `L` (side length of the square layer), `N` (number of nodes), and `kernel_name` (name of the kernel to use, “constant”, “linear”, “exponential”, or “gaussian”). An optional argument, `kernel_params`, can be used to specify the parameters of the



**Figure 4.3:** Exemplary PDF (A), CDF (B) and connectivity pattern (C). A Gaussian kernel is used, and the layer and mask size is  $L^2 = 1$ . The sharp kink at  $D = 0.5$  in the PDF is due to the mask. A kink exists in the CDF as well, marked by the arrow.

chosen kernel. Without this argument, sensible default values are used. The master seed value can be set using the optional argument `msd`. The value of `msd` is used to seed the PRNG used to draw the uniform, random positions for the nodes. NEST's global PRNG and each of the per-process PRNGs are seeded with the values  $(\text{msd} + 1, \text{msd} + 2, \dots, \text{msd} + n_{\text{vp}} + 1)$ , where  $n_{\text{vp}}$  is the number of virtual processes (VPs). Thus, for independent results, `msd` should differ by at least  $2 + n_{\text{vp}}$  between each instantiation of the class. The position of the source node is  $(0, 0)$  by default, but it can be changed with the optional `source_pos` argument. The mask is always centered around the source node, regardless of its position.

After the test object is created, the KS test can be run using the `ks_test` method. The KS test statistic and  $p$ -value is returned. The PDF is calculated on-the-fly as the product of the relevant kernel and the RDF. The CDF is then found by numerical integration, using the `quad` function from the `scipy.integrate` module. For finite integration limits the function uses the Clenshaw-Curtis method of numerical integration.

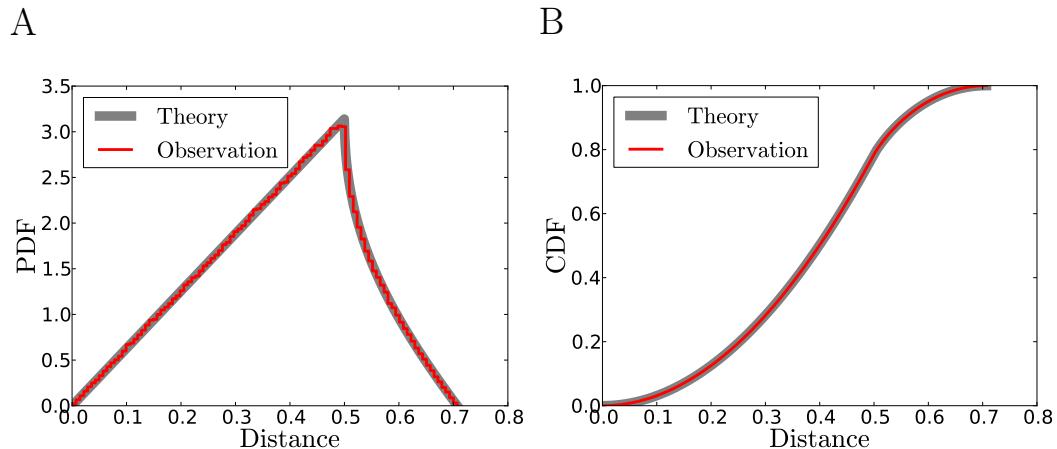
The method `z_test` implements the  $Z$ -test of the total connection count described earlier. The standard score and the two-sided  $p$ -value is returned.

### 4.1.2 Results

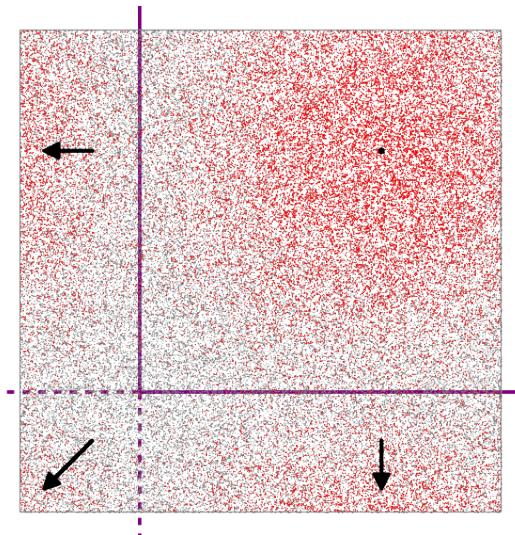
The test script in Appendix E will now be used to test the connection algorithms for connecting 2D layers of spatially structured networks in NEST.

With  $L = 1$ ,  $N = 1,000,000$ , and  $\text{msd} = 0$ , and with the constant kernel selected, the distribution of distances to connected nodes is as shown in Figure 4.4. There seems to be close agreement between the expected and observed distributions. According to the KS test, there is no evidence to reject the null hypothesis that the distances are drawn from the theoretical PDF ( $p = 0.321$ ). Similar results are found for the linear ( $p = 0.837$ ), exponential ( $p = 0.630$ ) and Gaussian ( $p = 0.852$ ) kernels.

By positioning the source node away from the center, while leaving the mask centered around the node, with the same  $L \times L$  size, the periodic boundary conditions come into play. The distribution of source-target distances should not change. An example can be seen in Figure 4.5. With the same parameters as above, but with the source node located at  $(L/4, L/4)$ , the  $p$ -values become 0.253, 0.091, 0.004, and 0.113 for constant, linear, exponen-



**Figure 4.4:** PDF (A) and CDF (B) of source-target distances, with layer size  $L^2 = 1$ , using a constant kernel. The grey lines show the theoretical predictions, and the red lines show the empirical observations. The empirical PDF is plotted with 100 bins.



**Figure 4.5:** Connectivity pattern with periodic boundary conditions and with the source node located at  $(L/4, L/4)$ . The purple line indicates the mask, and the arrows indicate the direction of shortest distance to the source node. The distribution of distances from the source node to connected nodes is the same as if the source node was located in the center.

tial, and Gaussian kernel, respectively. The  $p$ -value for the exponential kernel,  $p = 0.004$ , is suspiciously low. It appears, however, to be a statistical fluke. Rerunning the test 100 times with different master seed values results in  $p$ -values seemingly uniformly distributed on  $(0, 1)$ . A KS test of uniformity returns a  $p$ -values of 0.365.

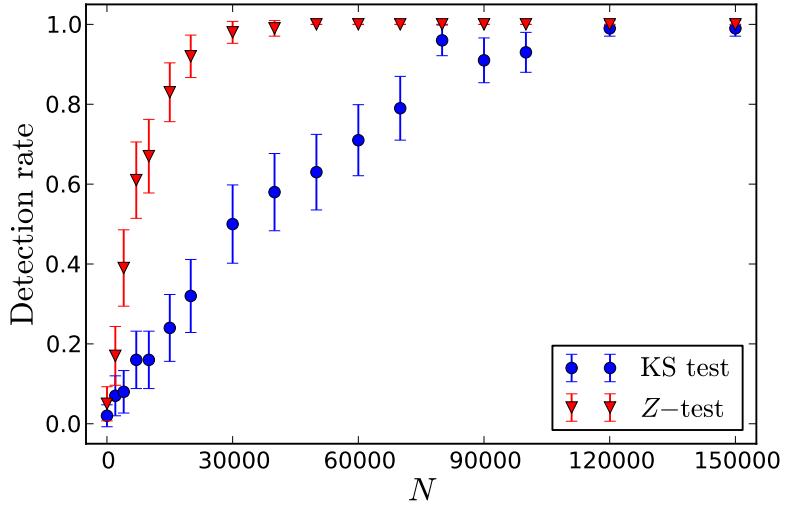
The tests can also be run with NEST using multiple VPs. This can be done by instantiating the test class with an extra argument `threads`. With 4 threads, there is no evidence of deviations from the expected distributions, both with constant ( $p = 0.321$ ), linear ( $p = 0.977$ ), exponential ( $p = 0.614$ ), and Gaussian ( $p = 0.899$ ) kernel.

## Sensitivity

As before, a control algorithm is implemented. It performs a Bernoulli trial on each node, with a probability  $p = \mathcal{P}(D)$ , given by the kernel, of success (a connection being made). By supplying the `ks_test` method (or the `z_test` method) with an extra argument `control=True`, NEST's connection algorithm is swapped with this control algorithm.

The control algorithm is used to examine the sensitivity of the test to errors and biases deliberately introduced into the data, as well as what effect different parameters have on the sensitivity. Note again that the reported  $p$ -values are only examples; a re-run with a different PRNG seed value will give a different  $p$ -value.

One bias was introduced by adding a constant  $c = 0.01$  to the Gaussian



**Figure 4.6:** Detection rate, e.g., proportion of tests that detect a problem with a bias deliberately introduced into the algorithm, as a function of  $N$ , for the KS test (blue) and the  $Z$ -test (red). Error bars show 90% confidence intervals. For this particular bias, the  $Z$ -test has a higher detection rate than the KS test.

kernel. This bias was easily detected with  $N = 1,000,000$  ( $p = 2.50 \times 10^{-29}$ ), as well as with  $N = 100,000$  ( $p = 6.43 \times 10^{-5}$ ). With  $N = 10,000$ , the bias was detected at significance level  $\alpha = 0.05$  ( $p = 0.0159$ ).

A second bias was introduced by increasing the distances passed to the Gaussian kernel by 1%. This was detected with  $N = 1,000,000$  ( $p = 1.59 \times 10^{-5}$ ), as well as with  $N = 100,000$  ( $p = 8.23 \times 10^{-3}$ ), but not with  $N = 10,000$  ( $p = 0.585$ ).

A third bias was introduced by excluding 1% of the nodes, randomly chosen, as potential targets. This bias is not detected by the KS test with any  $N$ , because there is no change in the overall distribution of connections. The  $Z$ -test, however, detects the bias, both with  $N = 1,000,000$  ( $p = 6.66 \times 10^{-16}$ ) and with  $N = 100,000$  ( $p = 5.32 \times 10^{-4}$ ), though not with  $N = 10,000$  ( $p = 0.988$ ).

As noted earlier, the  $p$ -values reported above are only examples. To get a sense for how consistently biases are detected, and how the sensitivity varies with  $N$ , we again introduce the first bias into the algorithm, i.e., we add a constant  $c = 0.01$  to the Gaussian kernel. We then choose the level of significance  $\alpha = 0.05$ , and consider  $p$ -values below  $\alpha$  to be a detection. Running the tests 100 times, each time with a different seed, the detection rate, i.e., proportion of times the bias is detected, can be found. In Figure 4.6, this detection rate is plotted as a function of the number of nodes  $N$ . Both tests consistently and reliably detect the bias for large  $N$ , though the sensitivity falls off rapidly with decreasing  $N$ . It is worth noting that the  $Z$ -test is more sensitive to this particular bias. Other biases, though, that change the distribution of distances, without affecting the overall connection count, are best detected by the KS

test.

Generally, both tests are quite sensitive to a range of different errors and biases, as long as  $N$  is large ( $\gtrsim 100,000$ ). Biases that do not change the distribution, but do affect the total number of connections, are not detected by the KS test, but such biases are detected quite consistently by the  $Z$ -test, as long as  $N$  is large.

## 4.2 Three-dimensional space

The procedure for testing the connectivity pattern of spatially structured networks in three-dimensional space is very similar to that for two-dimensional space, but the expressions for the density of potential targets at a given distance are different.

Let one node be centered on a cubic  $L \times L \times L$  layer. It can connect to  $N$  other nodes, uniformly distributed on the three-dimensional layer, with a cubic  $L \times L \times L$  mask, and a connection probability determined by a kernel  $\mathcal{P}(D)$ . The distance from node  $i$  to node  $j$  is

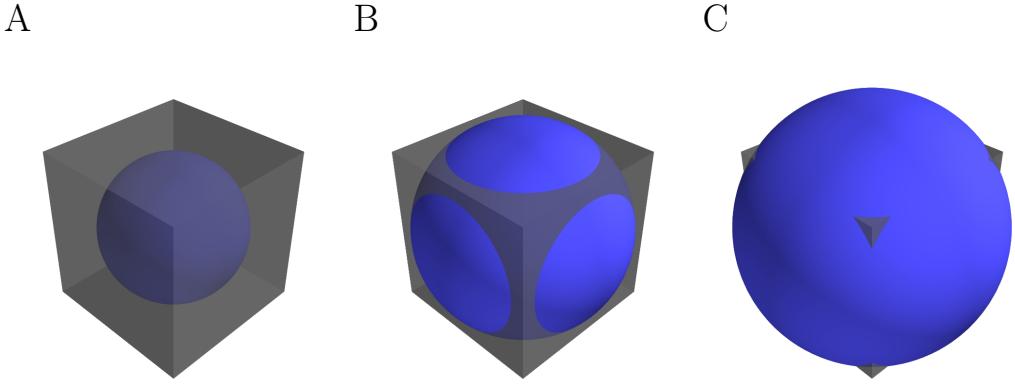
$$D_{ij} = \sqrt{\sum_{m=1}^k \Delta x_{ij}^2}, \quad (4.13)$$

where  $k$  now is three. With periodic boundary conditions,

$$\Delta x_{ij} = \begin{cases} |(x_i)_m - (x_j)_m| & \text{for } |(x_i)_m - (x_j)_m| \leq L/2 \\ L - |(x_i)_m - (x_j)_m| & \text{for } |(x_i)_m - (x_j)_m| > L/2 \end{cases} \quad (4.14)$$

The average node density is  $\rho_0 = N/L^3$ . The radial distribution function (RDF)  $\rho(D)$  is proportional to the surface area of a sphere with radius  $D$ . Thus, at a distance  $D \in [0, L/2]$ , the RDF is  $\rho(D) = \rho_0 4\pi D^2$ . For  $D \in (L/2, L/\sqrt{2}]$ , part of the sphere is outside the cubic mask. Specifically, a spherical cap will stick out of each of the six sides of the cube, as seen in Figure 4.7B. The surface area of each spherical cap is  $2\pi Dh$ , where  $h = D - L/2$  is the height of the cap. Subtracting the surface area of these caps from the total surface area of the sphere, we get  $4\pi D^2 - 6(2\pi Dh) = 2\pi D(3L - 4D)$ , and the RDF becomes  $\rho(D) = \rho_0 2\pi D(3L - 4D)$ . For  $D \in (L/\sqrt{2}, L\sqrt{3}/2]$ , the derivation becomes somewhat involved. The six spherical caps have “grown” to overlap each other near each of the 12 edges of the cube. This situation is seen in Figure 4.7C. If we subtract the surface area  $C$  of the spherical caps from the total area  $A$ , this overlap  $E$  has been subtracted twice, and has to be added again. Calling the surface area of the sphere inside the cube  $I$ , we can write

$$I = A - 6C + 12E, \quad (4.15)$$



**Figure 4.7:** **A:** For distances  $D \in [0, L/2]$ , the entire sphere is inside the cubic mask. **B:** For  $D \in (L/2, L/\sqrt{2}]$ , only part of the sphere is inside the mask, while six spherical caps are outside. **C:** For  $D \in (L/\sqrt{2}, L\sqrt{3}/2]$ , the six caps have “grown” to overlap each other.

where  $A = 4\pi D^2$  and  $C = 2\pi D(D - L/2)$ . To find  $E$ , consider Figure 4.8. The circle outlines one of the spherical caps, sitting on top of one of the cube’s sides. The sphere intersects one of the cube’s edges at points P and Q. A great circle passing through P and Q bisects the area  $E$  into two equal halves. Two more great circles pass through the center point R on the cap’s surface and P and Q, respectively. The spherical triangle PQR is further bisected into two equal spherical triangles by a fourth great circle passing through R. One such half (blue triangle in the figure) has a surface area  $T = D^2(\alpha + \beta + \gamma - \pi)$ , where  $\alpha$ ,  $\beta$ , and  $\gamma$  are the corner angles. The area  $\frac{1}{2}E$  can now be found by subtracting  $2T$  from the fraction  $2\gamma/2\pi$  of the spherical cap’s surface area  $C$  which is bounded by the two great circle arcs RP and RQ, i.e.,  $E = 2((2\gamma/2\pi)C - 2T)$ . The area  $I$  of the sphere inside the cube is therefore

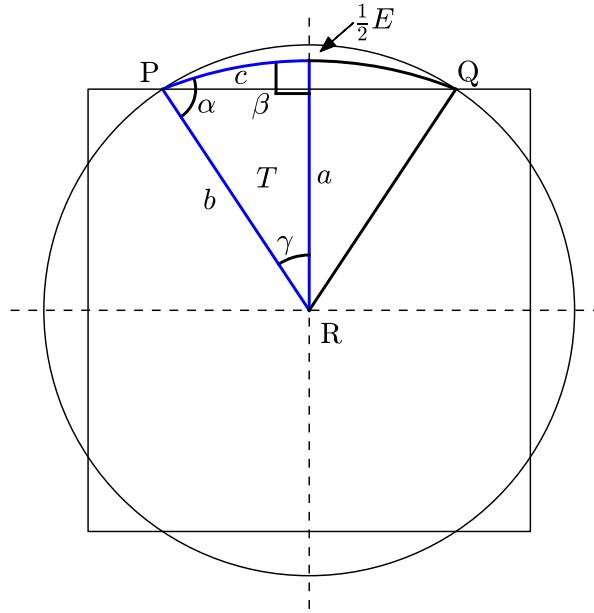
$$I = A - 6C + 12E \quad (4.16)$$

$$= A - 6C + 12\left(2\left(\frac{2\gamma}{2\pi}C - 2T\right)\right) \quad (4.17)$$

$$= A - 6C\left(1 - \frac{\gamma}{\pi}\right) - 48T. \quad (4.18)$$

Summarizing, the RDF is

$$\rho(D) = \begin{cases} \rho_0 A & \text{for } 0 \leq D \leq \frac{L}{2} \\ \rho_0 (A - 6C) & \text{for } \frac{L}{2} < D \leq \frac{L}{\sqrt{2}} \\ \rho_0 (A - 6C\left(1 - \frac{\gamma}{\pi}\right) - 48T) & \text{for } \frac{L}{\sqrt{2}} < D \leq \frac{L\sqrt{3}}{2} \\ 0 & \text{otherwise,} \end{cases} \quad (4.19)$$



**Figure 4.8:** Illustration of the derivation of the surface area  $E$  of the intersection between adjacent spherical caps. The square is one of the cube's sides, and the circle outlines one of the spherical caps, a portion of the sphere bounded by the extended plane of the cube's side. The area  $E$  is seen on the top of the figure, bisected by a great circle passing through points P and Q. The lines RP and RQ are also arcs of great circles. The surface area PQR is  $2T$ . The area  $\frac{1}{2}E$  can be found by subtracting  $2T$  from the fraction  $\frac{2\gamma}{2\pi}$  of the surface area of the spherical cap bounded by RP and RQ.

where

$$A = 4\pi D^2 \quad (4.20)$$

$$C = 2\pi D \left( D - \frac{L}{2} \right) \quad (4.21)$$

$$T = D^2 (\alpha + \beta + \gamma - \pi) \quad (4.22)$$

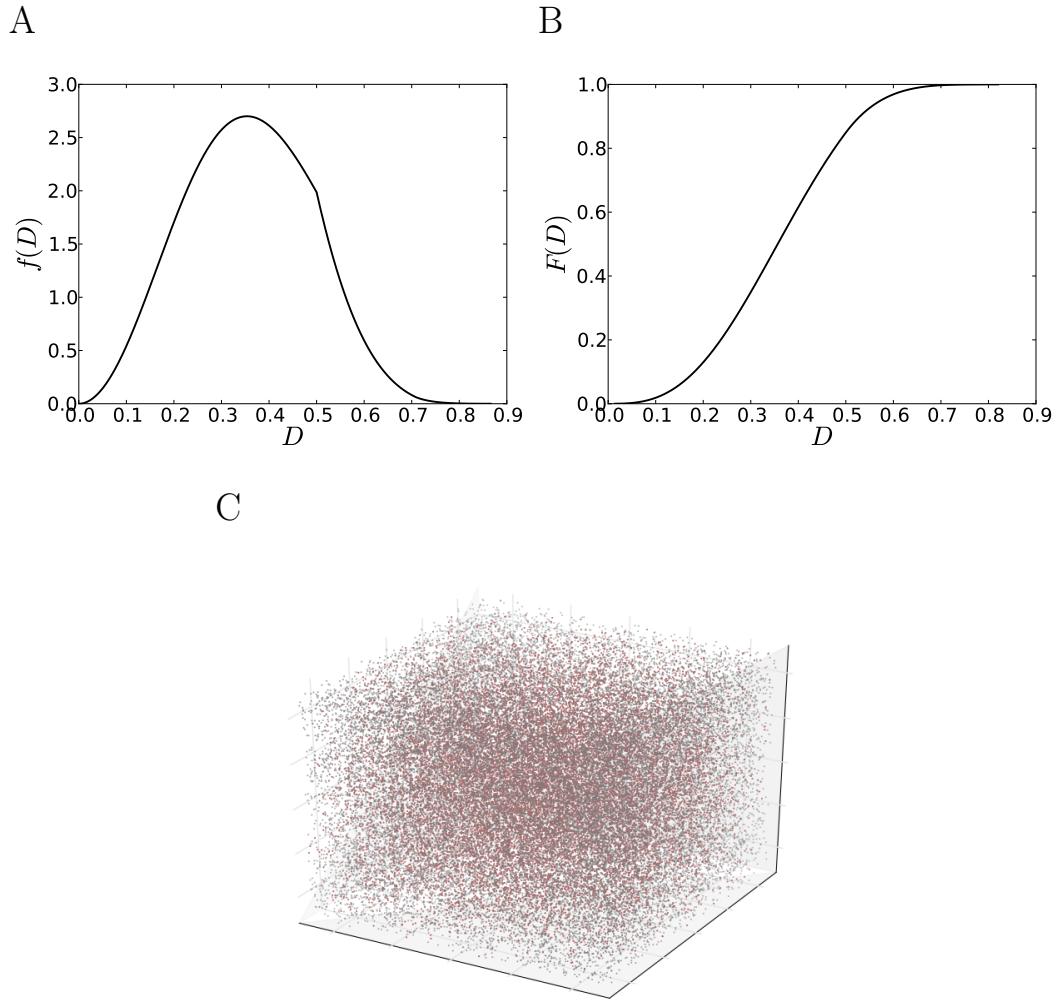
$$\alpha = \sin^{-1} \left( 1 / \sqrt{2 - \frac{L^2}{2D^2}} \right) \quad (4.23)$$

$$\beta = \frac{\pi}{2} \quad (4.24)$$

$$\gamma = \sin^{-1} \left( \sqrt{(1 - \frac{L^2}{2D^2}) / (1 - \frac{L^2}{4D^2})} \right). \quad (4.25)$$

As before, the PDF  $f(D)$  of source-target distances is the normalized product of the RDF  $\rho(D)$  and some kernel  $\mathcal{P}(D)$ , and the CDF  $F(D)$  can be found by numerical integration of  $f(D)$ . Figure 4.9 shows an exemplary PDF and CDF, together with the corresponding connectivity pattern.

The KS test is used to compare this theoretical CDF with the observed EDF. A Z-test that compares the observed number of connections with the expected number is also implemented. It works the same way as the Z-test implemented for 2D layers.



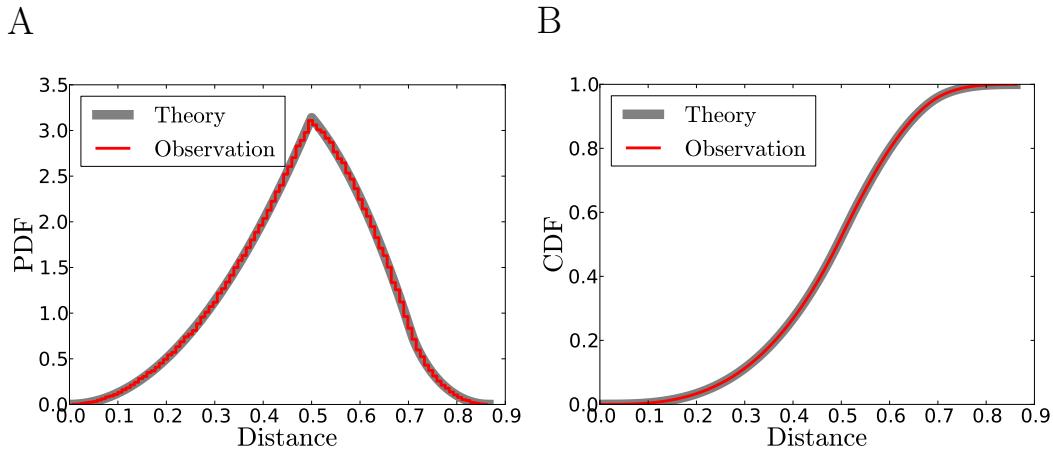
**Figure 4.9:** Exemplary PDF (A), CDF (B) and connectivity pattern (C) for a spatially structured network in 3D space. A Gaussian kernel is used, and the layer and mask size is  $L^3 = 1$ .

#### 4.2.1 Implementation

A script implementing the test procedure can be found in Appendix F. The usage is similar to that of 2D layers, and is demonstrated in the main section of the module. Unless `kernel_params` is specified, suitable defaults are used.

#### 4.2.2 Results

Using the script in Appendix F, NEST's connection algorithm for 3D spatially structured networks can be tested. With  $L = 1$ ,  $N = 1,000,000$ , there appears to be close agreement between the expected and the observed distribution of



**Figure 4.10:** PDF (A) and CDF (B) of source-target distances, with layer size  $L^3 = 1$ , using the constant kernel. The grey lines show the theoretical predictions, and the red lines show the empirical observations. The empirical PDF is plotted with 100 bins.

source-target distances for all four kernels (constant:  $p = 0.726$ , linear:  $p = 0.978$ , exponential:  $p = 0.803$ , and Gaussian:  $p = 0.770$ ). The theoretical and observed PDF and CDF for constant kernel are shown in Figure 4.10. When the source node is moved to  $(L/4, L/4, L/4)$  so that the periodic boundary conditions come into play, there is still a close agreement between theory and observation (constant:  $p = 0.965$ , linear:  $p = 0.367$ , exponential:  $p = 0.217$ , and Gaussian:  $p = 0.707$ ), and the same holds when the number of VPs is increased to 4 (constant:  $p = 0.726$ , linear:  $p = 0.600$ , exponential:  $p = 0.852$ , and Gaussian:  $p = 0.561$ ).

### Sensitivity

The same control algorithm that was implemented for 2D layers is implemented for 3D layers. It can be used both with the KS test and with the  $Z$ -test. We use it again to test the sensitivity of the test procedure to the same biases as for 2D layers. As before, we note that the reported  $p$ -values are only examples.

The first bias, introduced by having the Gaussian kernel function return a too high value, with a constant  $c = 0.01$  added, was detected by the KS test with  $N = 1,000,000$  ( $p = 1.52 \times 10^{-56}$ ) and  $N = 100,000$  ( $p = 1.48 \times 10^{-6}$ ), but not with  $N = 10,000$  ( $p = 0.350$ ). It was detected by the  $Z$ -test with  $N = 10,000$  ( $p = 6.25 \times 10^{-5}$ ) as well as with larger  $N$ , but not with  $N = 1,000$ .

The second bias, where the distances passed to the Gaussian kernel function are too high by 1%, is detected by the KS test with  $N = 1,000,000$  ( $p = 1.40 \times 10^{-9}$ ) and  $N = 100,000$  (at significance level  $\alpha = 0.05$ ;  $p = 0.019$ ), but not with  $N = 10,000$ . The  $Z$ -test also detects the bias with  $N = 100,000$  ( $p = 5.46 \times 10^{-8}$ ), but not with  $N = 10,000$  ( $p = 0.818$ ).

The third bias, where a randomly selected 1% of the nodes are excluded as potential targets, can not be detected by the KS test, regardless of  $N$ . The  $Z$ -test, however, detects the bias with  $N = 1,000,000$  ( $p = 3.38 \times 10^{-10}$ ),  $N = 100,000$  ( $p = 1.05 \times 10^{-4}$ ), and  $N = 10,000$  (at  $\alpha = 0.05$ ;  $p = 0.035$ ), but not with  $N = 1,000$  ( $p = 0.388$ ).

Overall, the KS test is sensitive when  $N$  is large, but the sensitivity falls off rapidly with decreasing  $N$ . Perhaps surprisingly, the  $Z$ -test is often more sensitive than the KS test, depending on the nature of the bias we wish to detect. The two tests are in some ways complementary to each other. The KS test detects biases that change the overall distribution of source-target distances, but it does not detect biases that affect the total number of connections without changing the distribution. The  $Z$ -test does detect biases that affect the total number of connections without changing the distribution.

### 4.3 Automated test procedure

We will now describe the automated test procedure for spatially structured networks, both in 2D and 3D space.

To maintain a high sensitivity while reducing the rate of false positives and the execution time, an adaptive testing strategy is used. First, a single KS test is run. If the resulting  $p$ -value is suspiciously low, a two-level test is run, comparing the output of  $n_{\text{runs}}$  KS tests to the expected uniform distribution, thereby either confirming or allaying our suspicion.

As discussed earlier, certain errors can not be detected by the KS test, but are detected by the  $Z$ -test. We therefore include the  $Z$ -test in the automated test procedure. The adaptive strategy can be used for the  $Z$ -test as well as for the KS test, re-running the  $Z$ -test  $n_{\text{runs}}$  times if the first test results in a suspicious  $p$ -value.

Let  $\alpha_1$  be the value below which the  $p$ -value from the single KS test (or  $Z$ -test) is considered suspicious, and  $\alpha_2$  be the value below which the  $p$ -value from the two-level test is considered too extreme, resulting in the rejection of  $H_0$ . The fraction of false positives will be  $\alpha = \alpha_1 \alpha_2$ . If  $k$  tests are included in the test suite, the number of false positives is binomially distributed. The probability of seeing one or more false positives is therefore

$$\begin{aligned} \text{Prob}(x \geq 1) &= 1 - \text{Prob}(x = 0) \\ &= 1 - \binom{k}{0} (\alpha_1 \alpha_2)^0 (1 - \alpha_1 \alpha_2)^k \\ &= 1 - (1 - \alpha_1 \alpha_2)^k \end{aligned}$$

The choice of  $\alpha_1$  and  $\alpha_2$  might thus depend on  $k$ , as well as other factors, such as the acceptable rate of false positives and the desired sensitivity and

the tests. The number of nodes  $N$  in the network and number of re-runs  $n_{\text{runs}}$  upon suspicion must also be chosen. Large values will clearly increase sensitivity, but the available computation time is a limiting factor.

### 4.3.1 Implementation

An implementation of the automated test procedure is found in Appendix G. Each of the four kernels (constant, linear, exponential, and Gaussian) are tested. In addition, tests are done with the source node shifted away from the center, and with multiple VPs. These six test configurations are tested both with the KS test and the  $Z$ -test, and both in two-dimensional and three-dimensional space. A total of 24 tests is thus run. With  $\alpha_1 = \alpha_2 = 0.01$ , the entire test suite will falsely report a problem with a probability  $1 - (1 - 0.01 \times 0.01)^{24} = 2.40 \times 10^{-3}$ . With  $N = 100,000$  nodes in the network and  $n_{\text{runs}} = 100$  re-runs upon suspicion, the test suite will in most cases complete in a couple of minutes.



# Chapter 5

## Discussion

We have developed tests for two main types of connection patterns, namely, random convergent and divergent connections with multapses and autapses allowed, and structured networks in two- and three-dimensional space with distance-dependent connection probability. Both tests have been implemented as Python test suites and have been used to test the connection routines in NEST. We emphasize that the test suites can be adapted to work with other simulators, simply by changing the function calls for generating the network and retrieving the resulting connections.

For random connections, a two-level test was proposed. Pearson's chi-squared test is used to test whether nodes are selected randomly and with equal probability. The resulting  $p$ -values are then compared with the expected uniform distribution using the Kolmogorov-Smirnov (KS) test. Advantages of this approach are increased sensitivity and the ability to detect too good fits as well as poor fits.

For spatially structured networks, expressions for the radial distribution function were derived, both for two- and three-dimensional space. The observed distribution of source-target distances could then be compared with the expected distribution, found as the normalized product of the radial distribution function  $\rho(D)$  and the distance-dependent connection probability (kernel)  $\mathcal{P}(D)$ , using the KS test.

We demonstrated the utility of the tests, both that they are able to detect a range of errors, and that they do not fail more often than expected when there is no error in the algorithms tested. To actually show this, rather than assume it is so, is important, because assumptions used when developing the tests might not be as accurate as assumed. For example, in the case of Pearson's chi-squared goodness-of-fit test, the assumption of uniformly distributed  $p$ -values is not strictly true, due to the discreteness of the  $p$ -values. And indeed, for very sparse data, the two-level test procedure used for testing random convergent or divergent connections was shown to not be reliable. It was demonstrated, however, that both tests (1) detect many deliberately introduced errors, and

(2) do not report more than the expected fraction of false positives under  $H_0$  (as long as data is not too sparse).

The tests were used on NEST’s probabilistic connection algorithms, with a range of different parameters, under different conditions. No evidence of errors or biases was found. For example, three-dimensional spatially structured networks with 1,000,000 nodes were created using each of the four kernels (constant, linear, exponential, and Gaussian), and the distribution of distances between connected nodes was tested using the KS test, resulting in the  $p$ -values 0.726, 0.978, 0.803, and 0.770, consistent with our expectations. This of course does not guarantee that no bias exists. Very small biases, or biases of a kind not easily detectable by the tests (e.g., biases caused by patterns in the underlying PRNG), might, indeed probably do, exist. Still, our confidence in the simulated connection patterns, and therefore the scientific findings based them, has grown.

An additional goal of this work was to develop automated test suites. An adaptive test strategy was proposed as a solution to the extra challenges this entails. A single test is first done on the algorithm being tested. If the resulting  $p$ -value is deemed suspicious, a two-level test is performed, comparing the  $p$ -values from tests of multiple network realizations with the expected uniform distribution. Using this strategy, the automated test suites achieve a low rate of false positives, fast run time, and a fairly high sensitivity. In certain cases one might want to opt for a safer alternative and do a small number of initial tests, instead of one, to determine whether more tests should be run. This will increase the sensitivity, while run time and rate of false positives will increase.

### Perspectives for future research

Variants of the probabilistic network types tested in this work exist, and tests have yet to be developed for these. The test for random convergent or divergent connection routines developed here, for example, assumes that autapses and multapses are allowed, and cannot be used when these are disallowed. Networks with disallowed autapses are relatively straightforward to test. We can simply make sure no node is connected to itself, and do a chi-squared test with the number of nodes available reduced by one. For disallowed multapses, however, an altogether different distribution will result.

For networks with spatial structure, tests for networks with a prescribed in- or out-degree  $C$ , as well as a kernel, remain to be developed. It is not entirely clear what is expected from such a connection routine, as a conflict between the two rules occur. In some cases, the value of the kernel is in this case interpreted not as a connection probability, but instead as relative probabilities ([Plessner and Austvoll 2009](#)). With kernel value for node  $i$  equal to  $k_i$ , the relative probability of connecting to node  $i$  is  $p_i = k_i/K$ , where  $K = \sum_i k_i$ . For such a connection routine, one possible test strategy is to use the chi-squared test,

with  $p_i C$  as the expectations.

Another case worth testing is spatially structured networks with open boundary conditions. When the source node is arbitrarily positioned, the boundary effects will in this case lead to somewhat involved expressions for the radial distribution of nodes.



# Appendix A

## NEST - A short tutorial

NEST (NEural Simulation Tool, [Gewaltig and Diesmann \(2007\)](#)) is a simulation environment developed by the NEST Initiative, capable of running simulations of large networks of point neurons. NEST is open source and can be downloaded free of charge from [nest-initiative.org](#). A brief introduction is given here. This section is based on the NEST tutorial by [Gewaltig, Morrison, and Plesser \(2012\)](#) and the NEST Topology User Manual by [Plesser and Enger \(2012\)](#). It mainly describes the functions implementing the connection algorithms tested in this thesis. A Python interface, PyNEST ([Eppler et al. 2009](#)), is available, and will be used here. Version 2.2.0 of NEST is used. To import the PyNEST module into Python we use the following command.

```
import nest
```

The concepts of *nodes* and *connections* described earlier are used in NEST. **Nodes** can be neurons, devices, or sub-networks. The neurons are either point-neurons (neurons with a single compartment) or neurons with a small number of compartments. They can be based on one of the many built-in neuron models, some of which are listed in Table A.1. The default parameters of each of these models can be modified. Due to NEST’s modular architecture, researchers can also create their own models from scratch.

Devices are nodes used to either stimulate or measure the activity of neurons. Some of the available devices are listed in Table A.2. Sub-networks (or subnets) are nodes who themselves are comprised of multiple nodes. Subnets can also be created inside other subnets. Using nested subnets we can create complex hierarchical structures of neurons with as many levels as we want.

The Python code below will create a neuron of type “iaf\_neuron”, using the default parameters for the model.

```
neuron1 = nest.Create('iaf_neuron')
```

Every node created is assigned number, called a global identifier (GID). This number is returned and assigned to the variable called `neuron1`. To create a

Name	Description
aEIF_cond_alpha	Conductance based exponential integrate-and-fire neuron model according to <a href="#">Brette and Gerstner (2005)</a> .
ginzburg_neuron	Binary stochastic neuron with sigmoidal activation function.
hh_psc_alpha	Spiking neuron using the Hodkin-Huxley formalism.
ht_neuron	Neuron model after <a href="#">Hill and Tononi (2005)</a> .
iaf_cond_alpha	Simple conductance based leaky integrate-and-fire neuron model. Post-synaptic change of conductance modelled by an alpha function.
iaf_cond_exp	Simple conductance based leaky integrate-and-fire neuron model. Post-synaptic change of conductance modeled by an exponential function.
iaf_neuron	Leaky integrate-and-fire model with alpha-function shaped synaptic currents.
izhikevich	Implementation of the simple spiking neuron model introduced by <a href="#">Izhikevich (2003)</a> .
mat2_psc_exp	Non-resetting leaky integrate-and-fire neuron model with exponential PSCs and adaptive threshold.
parrot_neuron	Neuron that repeats incoming spikes.
pp_psc_delta	Point process neuron with leaky integration of delta-shaped PSCs.
sli_neuron	The sli_neuron is a model whose state, update and calibration can be defined in SLI.

**Table A.1:** A few of the neuron models available in NEST.

neuron with some non-default parameters, say, the threshold potential  $V_{\text{th}}$  and the reset potential  $V_{\text{reset}}$ , the code will look as follows.

```
neuron2 = nest.Create('iaf_neuron',
                      params={'V_th': -50.0, 'V_reset': -65.0})
```

To create many identical nodes, the desired number of nodes is passed to `Create` as the second argument. In the following code 1,000 neurons are created, and the GIDs for all the neurons are stored as a list named `neurons`.

```
neurons = nest.Create('iaf_neuron', 1000)
```

**Connections** in NEST can represent synapses between neurons, but can also be connections from neurons to other types of nodes (devices and subnets). All connections are *directed*, *weighted* and *delayed*. Directed means information

Name	Description
ac_generator	This device produce an ac-current.
correlation_detector	Device for evaluating cross correlation between two spike sources
dc_generator	The DC-Generator provides a constant DC input to the connected node.
gamma_sup_generator	Simulate the superimposed spike train of a population of gamma process.
multimeter	Device to record analog data from neurons.
noise_generator	Device to generate Gaussian white noise current.
poisson_generator	Simulate neuron firing with Poisson processes statistics.
spike_detector	Device for detecting single spikes.
spike_generator	A device which generates spikes from an array with spike-times.
voltmeter	Device to record membrane potentials from neurons.

**Table A.2:** Some of the devices available in NEST.

travels in one direction. Weighted means the strength of connections can be varied. Delayed means it takes time for information to travel from one node to a connected node.

To manually connect one pair of nodes using the default synapse model “static\_synapse” with the default weight of 1.0 and the default delay of 1.0 milliseconds we can use the `Connect` function:

```
nest.Connect(neuron1, neuron2)
```

To use a weight of  $-1.5$  (negative weights means the connection will be inhibitory), a delay of 0.5 ms, and the synapse model “tsodyks\_synapse”, we pass some extra arguments to `Connect` as follows.

```
nest.Connect(n1, n2, -1.5, 0.5, model='tsodyks_synapse')
```

Table A.3 lists some of the synapse models in NEST. Several synapse models with short-term and long-term plasticity are available.

`Connect` can also be used to create many one-to-one connections between two lists of nodes. However, if we wish to make connections that are more complex than simple one-to-one connections, such as convergent and divergent connections, several other functions are available. One example is `RandomConvergentConnect`. The code snippet below first defines three variables  $N_s$ ,  $N_t$  and  $C$ . It then creates  $N_s$  neurons, which will serve

Name	Description
cont_delay_synapse	Synapse type for continuous delays.
ht_synapse	Synapse with depression after <a href="#">Hill and Tononi (2005)</a> .
static_synapse	Synapse type for static connections.
stdp_dopamine_synapse	Synapse type for dopamine-modulated spike-timing dependent plasticity.
stdp_synapse	Synapse type for spike-timing dependent plasticity.
tsodyks_synapse	Synapse type with short term plasticity.

**Table A.3:** Some of the synapse models available in NEST.

as *source* neurons, and  $N_t$  *target* neurons. It then connects them using `RandomConvergentConnect`.

```
N_s = 10
N_t = 10
C = 3
source_neurons = nest.Create('iaf_neuron', N_s)
target_neurons = nest.Create('iaf_neuron', N_t)
nest.RandomConvergentConnect(source_neurons, target_neurons, C)
```

The function call

```
nest.GetConnections(source_neurons)
```

will return the resulting connections. `RandomConvergentConnect` considers each node in the second list (`target_neurons`) in turn, and connects  $C$  randomly chosen nodes from the first list (`source_neurons`). Hence, the target nodes will all have an in-degree (number of incoming connections) equal to  $C$ , whereas the source nodes might have different out-degrees. Note that two or more connections (multapses) can be drawn between the same pair of nodes.

A similar function, `RandomDivergentConnect`, works in very much the same way, but it iterates over the *source* nodes instead of the target nodes, and draws random nodes from the *targets* instead of the sources. As a result, the source nodes will now all have the same out-degree  $C$ , while the in-degrees of the target nodes will vary. An example of a random divergent network was shown in Figure 1.1.

After having created all nodes and connected them, the simulation can be started with the `Simulate` function. `Simulate` only takes one argument; the number of milliseconds to simulate. The following example, borrowed from [Gewaltig, Morrison, and Plesser \(2012\)](#), is a simple simulation script. A neuron

receives stimuli (events) from one AC generator and two poisson generators. A voltmeter is also connected to the neuron to record the membrane potential. The simulation is run for 1,000 ms, and a plot of the membrane potential as a function of time is displayed.

```
import nest
import nest.voltage_trace
neuron = nest.Create('iaf_neuron')
sine = nest.Create('ac_generator', 1,
                   {'amplitude': 100.0, 'frequency': 2.0})
noise = nest.Create('poisson_generator', 2,
                     [{'rate': 70000.0}, {'rate': 20000.0}])
voltmeter = nest.Create('voltmeter', 1, {'withgid': True})
nest.Connect(sine, neuron)
nest.Connect(voltmeter, neuron)
nest.ConvergentConnect(noise, neuron, [1.0, -1.0], 1.0)
nest.Simulate(1000.0)
nest.voltage_trace.from_device(voltmeter)
```

## Spatially structured networks and the NEST Topology module

The simulation of spatially structured networks is increasingly a popular tool. A spatially structured network is a network where connection probabilities and -properties are determined by the spatial position of the nodes. With NEST, such networks can be created easily using the NEST Topology module ([Plessner and Austvoll 2009](#); [Plessner and Enger 2012](#)).

To use NEST Topology, the module has to be imported as show below. For convenience it is given the shorthand name “topo”.

```
import nest
import nest.topology as topo
```

In Topology, nodes are placed on *layers*. Layers are similar to subnets, but with extra information about the spatial position of each node. The term “layer” often refers to a two-dimensional structure, but layers in NEST Topology can be both two- and three-dimensional. They are either grid-based or free. In *grid-based layers*, nodes are placed on a Cartesian grid, while on *free layers*, nodes can be placed at arbitrary locations. The code below will create a simple two-dimensional grid-based layer.

```
layer_specs = {'elements': 'iaf_neuron',
               'rows': 10, 'columns': 10, 'extent': [2.0, 2.0]}
layer = topo.CreateLayer(layer_specs)
```

First a dictionary, `layer_specs`, is created, containing specifications for the layer. The node type is set to “iaf\_neuron”. The `rows` and `columns` entries specify the number of desired rows and columns of neurons, and `extent` is the size of the layer. Additional parameters can be included, such as the *x* and *y* position of the center of the layer (default is (0, 0)). The layer is assigned a GID, which is returned by `CreateLayer` and assigned to the variable `layer`.

To create a *free* layer, the position of every node must be passed to `CreateLayer`. If a list of positions is given, a node will be created at each. In the example below, three nodes are positioned on a two-dimensional layer.

```
pos = [[-0.3, 0.3], [0.4, 0.2], [0.0, -0.4]]
layer = topo.CreateLayer({'elements': 'iaf_neuron',
                         'positions': pos})
```

Here the extent is not specified, so the layer will have the default extent of  $1.0 \times 1.0$ .

To place a larger number of neurons at pseudorandom locations, we have to import and use a pseudorandom number generator (PRNG), such as the one supplied with `numpy`. Here, 1,000 neurons are uniformly distributed over the layer:

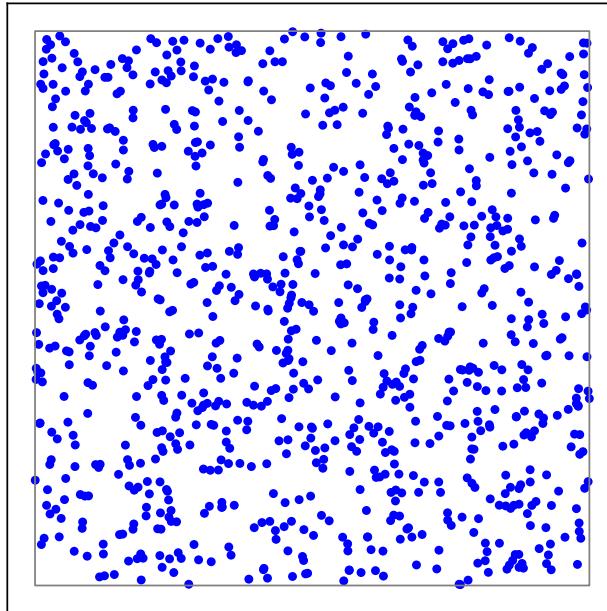
```
import numpy as np
x = np.random.uniform(-0.5, 0.5, 1000)
y = np.random.uniform(-0.5, 0.5, 1000)
pos = zip(x, y)
l = topo.CreateLayer({'elements': 'iaf_neuron',
                      'positions': pos})
topo.PlotLayer(l)
```

The last line will display a plot of all the nodes on the layer, similar to the one in Figure A.1. To create a three-dimensional layer we simply add a third component to the node positions.

```
import numpy as np
x = np.random.uniform(-0.5, 0.5, 1000)
y = np.random.uniform(-0.5, 0.5, 1000)
z = np.random.uniform(-0.5, 0.5, 1000)
pos = zip(x, y, z)
l = topo.CreateLayer({'elements': 'iaf_neuron',
                      'positions': pos})
topo.PlotLayer(l)
```

The result is seen in Figure A.2.

For small layers, a large fraction of the nodes will be close to the edge and therefore have fewer neighboring nodes. This might have an undesired effect on the simulation. To emulate the effect of a larger layer, the layers can



**Figure A.1:** A free layer with 1,000 nodes.  $x$  and  $y$  positions are drawn pseudorandomly from  $\mathcal{U}(-0.5, 0.5)$ .

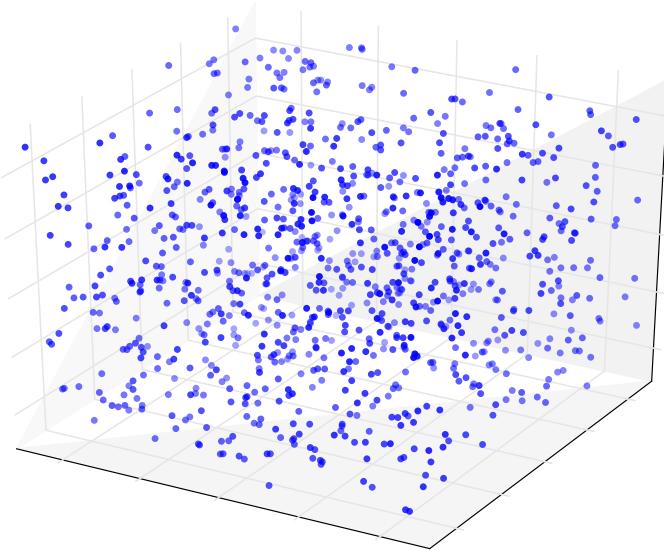
therefore be given periodic boundary conditions. The layer is effectively bent into a torus, so the left and right side are joined, and the top and bottom is joined.

To create a layer with periodic boundary conditions, an extra entry, '`edge_wrap': True`', is included in the dictionary passed to `CreateLayer`:

```
layer = topo.CreateLayer({'elements': 'iaf_neuron',
                         'positions': pos,
                         'edge_wrap': True})
```

The `ConnectLayers` function is used to connect two layers. One layer is considered the *source* layer, and the other the *target* layer. Connections between layers can be convergent or divergent. When creating a *convergent* connection, each node in the *target* layer is iterated over, and connections are drawn from the *source* layer. When creating a *divergent* connection, each node in the *source* layer is iterated over, and connections are drawn from the *target* layer. The layer which is iterated over is called the *driver*, and the layer from which nodes are drawn is called the *pool* layer.

When connecting two layers, we also specify a mask, and usually a kernel. A *mask* is a boundary around the considered node in the driver layer beyond which no nodes from the pool layer are considered. Masks can be rectangular, circular, or annulus shaped. The *kernel* is the distance-dependent connection probability function. Table A.4 lists all the available kernels in Topology. Users can also create their own kernels.



**Figure A.2:** A free three-dimensional layer with 1,000 nodes, where  $x$ ,  $y$ , and  $z$  positions are drawn pseudorandomly from  $\mathcal{U}(-0.5, 0.5)$ .

In the example below, two identical layers are created, and then connected using a Gaussian kernel. The result can be seen in Figure A.3.

```

1 import nest
2 import nest.topology as topo
3
4 layer_specs = {'elements': 'iaf_neuron',
5                 'rows': 21, 'columns': 21}
6 source_layer = topo.CreateLayer(layer_specs)
7 target_layer = topo.CreateLayer(layer_specs)
8
9 mask = {'rectangular': {'lower_left': [-0.4, -0.4],
10                         'upper_right': [0.4, 0.4]}}
11 kernel = {'gaussian': {'p_center': 1., 'sigma': .2}}
12 conn_specs = {'connection_type': 'convergent',
13                 'mask': mask, 'kernel': kernel}
14 topo.ConnectLayers(source_layer, target_layer, conn_specs)
15
16 fig = topo.PlotLayer(target_layer, nodesize = 60,
17                       nodecolor = 'grey')
18
19 center_node = topo.FindCenterElement(source_layer)
20 topo.PlotTargets(center_node, target_layer, fig = fig,
21                   mask = mask, kernel = kernel,
```

Name	Parameters	Function
constant		$p \in [0, 1]$
uniform	$min, max$	$p \in [min, max)$ uniformly
linear	$a, c$	$p(d) = c + ad$
exponential	$a, c, \tau$	$p(d) = c + ae^{-\frac{d}{\tau}}$
gaussian	$p_{center}, \sigma, \mu, c$	$p(d) = c + p_{center}e^{-\frac{(d-\mu)^2}{2\sigma^2}}$
gaussian2D	$p_c, \sigma_x, \sigma_y, \mu_x, \mu_y, \rho, c$	$p(d) = c + p_c e^{-\frac{\frac{(dx-\mu_x)^2}{\sigma_x^2} - \frac{(dy-\mu_y)^2}{\sigma_y^2} + 2\rho \frac{(dx-\mu_x)(dy-\mu_y)}{\sigma_x \sigma_y}}{2(1-\rho^2)}}$

**Table A.4:** Kernels available in NEST Topology.

```

22     src_size = 250, src_color = 'red',
23     tgt_size = 30, tgt_color = 'red',
24     kernel_color = 'blue')

```

In lines 4 and 5, layer specifications are stored in the variable `layer_specs`. No extent is specified, so the default extent of  $1.0 \times 1.0$  will be used. In line 6 and 7, these specs are used to create two identical layers, called `source_layer` and `target_layer`. In line 9 through 11, mask and kernel specifications are stored. A rectangular mask is chosen, and the extent of the mask is specified in terms of the locations of the lower left and upper right corners. A Gaussian kernel is chosen, and the two required parameters of the function, the connection probability at the center (`p_center`) and the standard deviation (`sigma`), are set. In line 12 and 13 the connection specifications are set using the already defined mask and kernel, and the connection type set to convergent. In line 14 the layers are actually connected. The rest of the code generates the plot seen in Figure A.3.

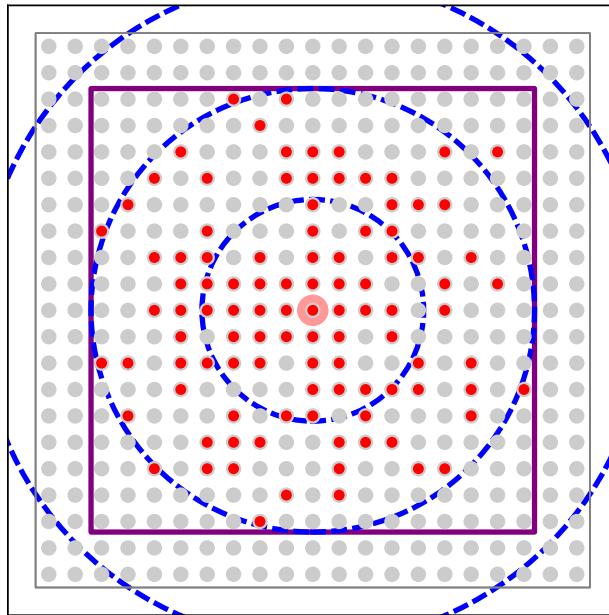
It is possible to specify a fixed degree for the driver layer, similar to what is done when using `RandomConvergentConnect` for networks without spatial structure. In this case, each node in the driver layer selects randomly chosen nodes from the pool layer and either connects or not based on the probability given by the kernel, until the number of connections matches the specified degree. This will allow multapses to be formed (unless specifically prohibited). To implement a fixed driver layer degree, only a small change to the parameters given to `ConnectLayers` is needed:

```

conn_specs = {'connection_type': 'convergent',
              'mask': mask, 'kernel': kernel,
              'number_of_connections': 100,
              'allow_multapses': True}

```

Several functions for inspecting different aspects of the network is available, in addition to the plotting functions demonstrated above. All the query



**Figure A.3:** Plot of a layer, with neurons connected to the source node in the center marked in red. The mask is shown in purple, and the blue rings mark  $\sigma$ ,  $2\sigma$  and  $3\sigma$  from the Gaussian kernel function. No nodes outside the mask is connected to the source node. As expected, the fraction of connected nodes is higher close to the source node. Note however, that the actual number of connected nodes is higher between the  $1\sigma$  mark and the  $2\sigma$  mark than between the center and the  $1\sigma$  mark. This is because a greater number of nodes are available further away from the center.

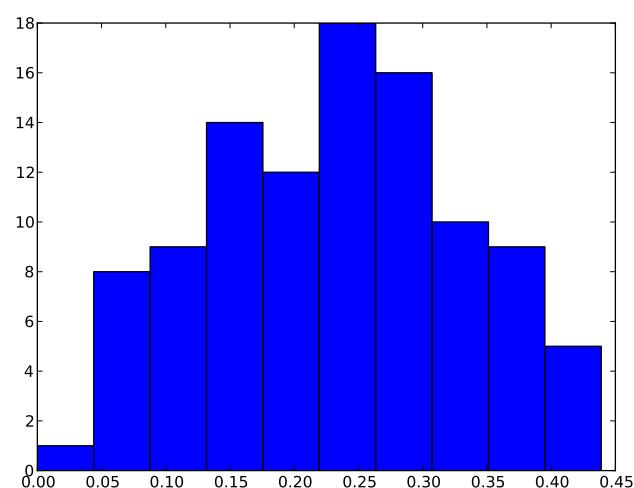
functions from the main nest module, such as the `GetTargetNodes` function explained earlier, can also be used on layered networks created with the Topology module. In addition, the topology module provides some extra query functions, such as `FindCenterElement`, `GetTargetPositions`, `Distance`, and `Displacement`. The usage of some of these are demonstrated in the code below, which is a continuation of the previous example.

```

25 targets = topo.GetTargetNodes(center_node, target_layer)[0]
26 distances = topo.Distance(center_node, targets)
27 import matplotlib.pyplot as plt
28 plt.figure()
29 plt.hist(distances, bins=10)

```

This will generate a histogram like the one in Figure A.4, showing the distribution of distances between the center source node and connected nodes in the target layer.



**Figure A.4:** Distribution of distances between a centered source node and the connected target nodes.



# Appendix B

## Test script for random convergent connections

```
1  ,,
2  @author: Daniel Hjertholm
3  Tests for RandomConvergentConnect .
4  ,,
5
6
7  import numpy
8  import numpy.random as rnd
9  import scipy.stats
10 import matplotlib.pyplot as plt
11 import nest
12
13
14 class RCC_tester(object):
15   ,,
16   Class used for testing RandomConvergentConnect.
17   ,,
18
19   def __init__(self, N_s, N_t, C, e_min=10, threads=1):
20   ,,
21   Sets up the experiment, and calculates expected distributions
22   for later comparison with the observed distribution.
23
24   Parameters
25   -----
26   N_s      : Number of source neurons.
27   N_t      : Number of target neurons.
28   C        : In-degree (number of connections per
29                  target neuron).
30   e_min    : Minimum expected number of observations in
31                  each bin. Default is 10.
32   threads  : Set number of local threads. Default is 1.
33   ,,
34
35   self.N_s = N_s
36   self.N_t = N_t
37   self.C = C
38   self.e_min = e_min
39   self.threads = threads
40
41   nest.ResetKernel()
42   nest.SetKernelStatus({'local_num_threads': self.threads})
43   self.n_vp = nest.GetKernelStatus('total_num_virtual_procs')
44
45   expected_degree = self.N_t * self.C / float(self.N_s)
46   if expected_degree < self.e_min:
47     raise RuntimeWarning(
48       'Expected out-degree (%.2f) is less than e_min (%.2f). , \
49       Increase N_t*C / N_s or decrease e_min. , % \
50       (expected_degree, self.e_min)')
51
52   self.expected = [expected_degree] * self.N_s
53
54   def _counter(self, x):
55     ,,
56     Count similar elements in list.
57
58   Parameters
59   -----
59   x: Any list.
60
61   Return values
62   -----
63   ,,
64   list containing counts of similar elements.
65   ,,
66
67   start = min(x)
68   counts = [0] * self.N_s
69   for elem in x:
70     counts[elem - start] += 1
71
72   return counts
73
74   def _reset(self, msd):
75     ,,
76     Reset the NEST kernel and set seed values.
77
78   Parameters
79   -----
79   msd: Master RNG seed.
80   ,,
81
82
83   nest.ResetKernel()
84   nest.SetKernelStatus({'local_num_threads': self.threads})
85   self.n_vp = nest.GetKernelStatus('total_num_virtual_procs')
86
87   # Set PRNG seed values:
88   if msd == None:
89     msd = rnd.randint(1000000)
```

```

90     msdrange = range(msd, msd + self.n_vp)
91     nest.SetKernelStatus({'rng_seed': msd + self.n_vp,
92                           'rng_seeds': msdrange})
93
94     def _build(self):
95         '''Create all nodes.'''
96
97         self.source_nodes = nest.Create('iaf_neuron', self.N_S)
98         self.target_nodes = nest.Create('iaf_neuron', self.N_T)
99
100    def _connect(self):
101        '''Connect all nodes.'''
102
103        nest.RandomConvergentConnect(self.source_nodes, self.target_nodes,
104                                      self.C, options={'allow_multapses': True})
105
106    def _get_degrees(self, msd):
107        '''
108            Reset NEST, create nodes, connect them, and retrieve the
109            resulting connections.
110
111            Parameters
112            -----
113            msd: Master RNG seed.
114
115            Return values
116            -----
117            list containing the out-degrees of the source nodes.
118
119            self._reset(msd)
120            self._build()
121            self._connect()
122
123        connections = nest.GetConnections(source=self.source_nodes)
124        source_connections = [conn[0] for conn in connections]
125        degrees = self._counter(source_connections)
126
127        return degrees
128
129    def _get_degrees_control(self, msd):
130        '''
131            Instead of using NEST, this method returns data with the expected
132            multinomial distribution.
133
134            Parameters
135            -----
136            msd: Master RNG seed.
137
138            Return values
139            -----
140            list containing the out-degrees of the source nodes.
141
142        ,,
143        if msd != None:
144            rnd.seed(msd)
145
146            con = rnd.randint(0, self.N_S, self.N_T * self.C)
147
148            degrees = self._counter(con)
149
150            return degrees
151
152    def chi_squared_test(self, msd=None, control=False):
153        '''
154            Create a single network and compare the resulting out-degree
155            distribution with the expected distribution using Pearson's chi-squared
156            GOF test.
157
158            Parameters
159            -----
160            msd : Master RNG seed.
161            control: Boolean value. If True, _get_degrees_control will
162            be used instead of _get_degrees.
163
164            Return values
165            -----
166            chi-squared statistic.
167            p-value from chi-squared test.
168
169            if control:
170                degrees = self._get_degrees_control(msd)
171            else:
172                degrees = self._get_degrees(msd)
173
174
175            # ddof: adjustment to the degrees of freedom. df = k-1-ddof
176            return scipy.stats.chisquare(numpy.array(degrees),
177                                         numpy.array(self.expected), ddof=0)
178
179    def two_level_test(self, n_runs, start_seed=None, control=False,
180                      show_histogram=False, histogram_bins=100,
181                      show_CDF=False):
182
183        '''
184            Create a network and run chi-squared GOF test n_runs times.
185            Test whether resulting p-values are uniformly distributed
186            on [0, 1] using the Kolmogorov-Smirnov GOF test.
187
188            Parameters
189            -----
190            n_runs : number of times to repeat chi-squared test.
191            start_seed : First PRNG seed value.
192            control : boolean value. If True, -get_degrees_control will be
193            used instead of -get_degrees.
194
195            show_histogram: Specify whether histogram should be displayed.
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
888
889
889
889
890
890
890
891
891
891
892
892
892
893
893
893
894
894
894
895
895
895
896
896
896
897
897
897
898
898
898
899
899
899
900
900
900
901
901
901
902
902
902
903
903
903
904
904
904
905
905
905
906
906
906
907
907
907
908
908
908
909
909
909
910
910
910
911
911
911
912
912
912
913
913
913
914
914
914
915
915
915
916
916
916
917
917
917
918
918
918
919
919
919
920
920
920
921
921
921
922
922
922
923
923
923
924
924
924
925
925
925
926
926
926
927
927
927
928
928
928
929
929
929
930
930
930
931
931
931
932
932
932
933
933
933
934
934
934
935
935
935
936
936
936
937
937
937
938
938
938
939
939
939
940
940
940
941
941
941
942
942
942
943
943
943
944
944
944
945
945
945
946
946
946
947
947
947
948
948
948
949
949
949
950
950
950
951
951
951
952
952
952
953
953
953
954
954
954
955
955
955
956
956
956
957
957
957
958
958
958
959
959
959
960
960
960
961
961
961
962
962
962
963
963
963
964
964
964
965
965
965
966
966
966
967
967
967
968
968
968
969
969
969
970
970
970
971
971
971
972
972
972
973
973
973
974
974
974
975
975
975
976
976
976
977
977
977
978
978
978
979
979
979
980
980
980
981
981
981
982
982
982
983
983
983
984
984
984
985
985
985
986
986
986
987
987
987
988
988
988
989
989
989
990
990
990
991
991
991
992
992
992
993
993
993
994
994
994
995
995
995
996
996
996
997
997
997
998
998
998
999
999
999
1000
1000
1000
1001
1001
1001
1002
1002
1002
1003
1003
1003
1004
1004
1004
1005
1005
1005
1006
1006
1006
1007
1007
1007
1008
1008
1008
1009
1009
1009
1010
1010
1010
1011
1011
1011
1012
1012
1012
1013
1013
1013
1014
1014
1014
1015
1015
1015
1016
1016
1016
1017
1017
1017
1018
1018
1018
1019
1019
1019
1020
1020
1020
1021
1021
1021
1022
1022
1022
1023
1023
1023
1024
1024
1024
1025
1025
1025
1026
1026
1026
1027
1027
1027
1028
1028
1028
1029
1029
1029
1030
1030
1030
1031
1031
1031
1032
1032
1032
1033
1033
1033
1034
1034
1034
1035
1035
1035
1036
1036
1036
1037
1037
1037
1038
1038
1038
1039
1039
1039
1040
1040
1040
1041
1041
1041
1042
1042
1042
1043
1043
1043
1044
1044
1044
1045
1045
1045
1046
1046
1046
1047
1047
1047
1048
1048
1048
1049
1049
1049
1050
1050
1050
1051
1051
1051
1052
1052
1052
1053
1053
1053
1054
1054
1054
1055
1055
1055
1056
1056
1056
1057
1057
1057
1058
1058
1058
1059
1059
1059
1060
1060
1060
1061
1061
1061
1062
1062
1062
1063
1063
1063
1064
1064
1064
1065
1065
1065
1066
1066
1066
1067
1067
1067
1068
1068
1068
1069
1069
1069
1070
1070
1070
1071
1071
1071
1072
1072
1072
1073
1073
1073
1074
1074
1074
1075
1075
1075
1076
1076
1076
1077
1077
1077
1078
1078
1078
1079
1079
1079
1080
1080
1080
1081
1081
1081
1082
1082
1082
1083
1083
1083
1084
1084
1084
1085
1085
1085
1086
1086
1086
1087
1087
1087
1088
1088
1088
1089
1089
1089
1090
1090
1090
1091
1091
1091
1092
1092
1092
1093
1093
1093
1094
1094
1094
1095
1095
1095
1096
1096
1096
1097
1097
1097
1098
1098
1098
1099
1099
1099
1100
1100
1100
1101
1101
1101
1102
1102
1102
1103
1103
1103
1104
1104
1104
1105
1105
1105
1106
1106
1106
1107
1107
1107
1108
1108
1108
1109
1109
1109
1110
1110
1110
1111
1111
1111
1112
1112
1112
1113
1113
1113
1114
1114
1114
1115
1115
1115
1116
1116
1116
1117
1117
1117
1118
1118
1118
1119
1119
1119
1120
1120
1120
1121
1121
1121
1122
1122
1122
1123
1123
1123
1124
1124
1124
1125
1125
1125
1126
1126
1126
1127
1127
1127
1128
1128
1128
1129
1129
1129
1130
1130
1130
1131
1131
1131
1132
1132
1132
1133
1133
1133
1134
1134
1134
1135
1135
1135
1136
1136
1136
1137
1137
1137
1138
1138
1138
1139
1139
1139
1140
1140
1140
1141
1141
1141
1142
1142
1142
1143
1143
1143
1144
1144
1144
1145
1145
1145
1146
1146
1146
1147
1147
1147
1148
1148
1148
1149
1149
1149
1150
1150
1150
1151
1151
1151
1152
1152
1152
1153
1153
1153
1154
1154
1154
1155
1155
1155
1156
1156
1156
1157
1157
1157
1158
1158
1158
1159
1159
1159
1160
1160
1160
1161
1161
1161
1162
1162
1162
1163
1163
1163
1164
1164
1164
1165
1165
1165
1166
1166
1166
1167
1167
1167
1168
1168
1168
1169
1169
1169
1170
1170
1170
1171
1171
1171
1172
1172
1172
1173
1173
1173
1174
1174
1174
1175
1175
1175
1176
1176
1176
1177
1177
1177
1178
1178
1178
1179
1179
1179
1180
1180
1180
1181
1181
1181
1182
1182
1182
1183
1183
1183
1184
1184
1184
1185
1185
1185
1186
1186
1186
1187
1187
1187
1188
1188
1188
1189
1189
1189
1190
1190
1190
1191
1191
1191
1192
1192
1192
1193
1193
1193
1194
1194
1194
1195
1195
1195
1196
1196
1196
1197
1197
1197
1198
1198
1198
1199
1199
1199
1200
1200
1200
1201
1201
1201
1202
1202
1202
1203
1203
1203
1204
1204
1204
1205
1205
1205
1206
1206
1206
1207
1207
1207
1208
1208
1208
1209
1209
1209
1210
1210
1210
1211
1211
1211
1212
1212
1212
1213
1213
1213
1214
1214
1214
1215
1215
1215
1216
1216
1216
1217
1217
1217
1218
1218
1218
1219
1219
1219
1220
1220
1220
1221
1221
1221
1222
1222
1222
1223
1223
1223
1224
1224
1224
1225
1225
1225
1226
1226
1226
1227
1227
1227
1228
1228
1228
1229
1229
1229
1230
1230
1230
1231
1231
1231
1232
1232
1232
1233
1233
1233
1234
1234
1234
1235
1235
1235
1236
1236
1236
1237
1237
1237
1238
1238
1238
1239
1239
1239
1240
1240
1240
1241
1241
1241
1242
1242
1242
1243
1243
1243
1244
1244
1244
1245
1245
1245
1246
1246
1246
1247
1247
1247
1248
1248
1248
1249
1249
1249
1250
1250
1250
1251
1251
1251
1252
1252
1252
1253
1253
1253
1254
1254
1254
1255
1255
1255
1256
1256
1256
1257
1257
1257
1258
1258
1258
1259
1259
1259
1260
1260
1260
1261
1261
1261
1262
1262
1262
1263
1263
1263
1264
1264
1264
1265
1265
1265
1266
1266
1266
1267
1267
1267
1268
1268
1268
1269
1269
1269
1270
1270
1270
1271
1271
1271
1272
1272
1272
1273
1273
1273
1274
1274
1274
1275
1275
1275
1276
1276
1276
1277
1277
1277
1278
1278
1278
1279
1279
1279
1280
1280
1280
1281
1281
1281
1282
1282
1282
1283
1283
1283
1284
1284
1284
1285
1285
1285
1286
1286
1286
1287
1287
1287
1288
1288
1288
1289
1289
1289
1290
1290
1290
1291
1291
1291
1292
1292
1292
1293
1293
1293
1294
1294
1294
1295
1295
1295
1296
1296
1296
1297
1297
1297
1298
1298
1298
1299
1299
1299
1300
1300
1300
1301
1301
1301
1302
1302
1302
1303
1303
1303
1304
1304
1304
1305
1305
1305
1306
1306
1306
1307
1307
1307
1308
1308
1308
1309
1309
1309
1310
1310
1310
1311
1311
1311
1312
1312
1312
1313
1313
1313
1314
1314
1314
1315
1315
1315
1316
1316
1316
1317
1317
1317
1318
1318
1318
1319
1319
1319
1320
1320
1320
1321
1321
1321
1322
1322
1322
1323
1323
1323
1324
1324
1324
1325
1325
1325
1326
1326
1326
1327
1327
1327
1328
1328
1328
1329
1329
1329
1330
1330
1330
1331
1331
1331
1332
1332
1332
1333
1333
1333
1334
1334
1334
1335
1335
1335
1336
1336
1336
1337
1337
1337
1338
1338
1338
1339
1339
1339
1340
1340
1340
1341
1341
1341
1342
1342
1342
1343
1343
1343
1344
1344
1344
1345
1345
1345
1346
1346
1346
1347
1347
1347
1348
1348
1348
1349
1349
1349
1350
1350
1350
1351
1351
1351
1352
1352
1352
1353
1353
1353
1354
1354
1354
1355
1355
1355
1356
1356
1356
1357
1357
1357
1358
1358
1358
1359
1359
1359
1360
1360
1360
1361
1361
1361
1362
1362
1362
1363
1363
1363
1364
1364
1364
1365
1365
1365
1366
1366
1366
1367
1367
1367
1368
1368
1368
1369
1369
1369
1370
1370
1370
1371
1371
1371
1372
1372
1372
1373
1373
1373
1374
1374
1374
1375
1375
1375
1376
1376
1376
1377
1377
1377
1378
1378
1378
1379
1379
1379
1380
1380
1380
1381
1381
1381
1382
1382
1382
1383
1383
1383
1384
1384
1384
1385
1385
1385
1386
1386
1386
1387
1387
1387
1388
1388
1388

```

```

194     histogram_bins: Number of histogram bins.
195     show_CDF : Specify whether EDF should be displayed.
196
197     Return values
198
199     -----
200     KS statistic.
201     p-value from KS test.
202
203     self.pvalues = []
204
205     if start_seed == None:
206         for i in range(n_runs):
207             print 'Running test %d, %d, %d, %d' % (i + 1, n_runs)
208             chi, p = self.chi_squared_test(None, control)
209             self.pvalues.append(p)
210
211     else:
212         seed_jump = self.n_vp + 1
213         end_seed = start_seed + n_runs * seed_jump
214         for seed in range(start_seed, end_seed, seed_jump):
215             print 'Running test %d, %d, %d' % \
216                 (1 + (seed - start_seed) / seed_jump, n_runs)
217             chi, p = self.chi_squared_test(seed, control)
218             self.pvalues.append(p)
219
220     ks, p = scipy.stats.ks_2samp(self.pvalues, 'uniform',
221                                 alternative='two-sided')
222
223     if show_CDF:
224         plt.figure()
225         self.pvalues.sort()
226         y = [(i + 1.) / len(self.pvalues)
227               for i in range(len(self.pvalues))]
228         plt.step([0.0] + self.pvalues + [1.0], [0.0] + y + [1.0])
229         plt.xlabel('P-values')
230         plt.ylabel('Empirical distribution function')
231
232     if show_histogram:
233         plt.figure()
234         plt.hist(self.pvalues, bins=histogram_bins)
235         plt.xlabel('P-values')
236         plt.ylabel('Frequency')
237
238     if show_CDF or show_histogram:
239         plt.show(block=True)
240
241     return ks, p
242
243 if __name__ == '__main__':
244     test = RCC_tester(N_s=1000, N_t=1000, C=1000)
245     ks, p = test.two_level_test(n_runs=1000, start_seed=0)

```



# Appendix C

## Test script for random divergent connections

```
1  ,,
2  @author: Daniel Hjertholm
3  Tests for RandomDivergentConnect.
4  ,,
5  ,
6
7  import numpy
8  import numpy.random as rnd
9  import scipy.stats
10 import matplotlib.pyplot as plt
11 import nest
12
13
14 class RDC_tester(object):
15   ,,
16   Class used for testing RandomDivergentConnect.
17   ,,
18
19   def __init__(self, N_s, N_t, C, e_min=10, threads=1):
20   ,,
21   Sets up the experiment, and calculates expected distributions
22   for later comparison with the observed distribution.
23
24   Parameters
25   -----
26   N_s      : Number of source neurons.
27   N_t      : Number of target neurons.
28   C        : Out-degree (number of connections per
29                  source neuron).
30   e_min    : Minimum expected number of observations in
31                  each bin. Default is 10.
32   threads: Set number of local threads. Default is 1.
33   ,,
34   self.N_s = N_s
35   self.N_t = N_t
36   self.C = C
37   self.e_min = e_min
38   self.threads = threads
39
40   nest.ResetKernel()
41   nest.SetKernelStatus({'local_num_threads': self.threads})
42   nest.SetKernelStatus({'total_num_virtual_procs': self.n_vp})
43   self.n_vp = nest.GetKernelStatus('total_num_virtual_procs')
44
45   expected_degree = self.N_s * self.C / float(self.N_t)
46   if expected_degree < self.e_min:
47     raise RuntimeWarning(
48       'Expected in-degree (%.2f) is less than e_min (%.2f). '\
49       ', Increase N_s*C / N_t or decrease e_min. '%
50       '(expected_degree, self.e_min)')
51
52   self.expected = [expected_degree] * self.N_t
53
54   def _counter(self, x):
55   ,,
56   Count similar elements in list.
57
58   Parameters
59   -----
59   x: Any list.
60
61   Return values
62   -----
63   list containing counts of similar elements.
64
65   ,,
66
67   start = min(x)
68   counts = [0] * self.N_s
69   for elem in x:
70     counts[elem - start] += 1
71
72   return counts
73
74   def _reset(self, msd):
75   ,,
76   Reset the NEST kernel and set seed values.
77
78   Parameters
79   -----
79   msd: Master RNG seed.
80   ,,
81
82   nest.ResetKernel()
83   nest.SetKernelStatus({'local_num_threads': self.threads})
84   self.n_vp = nest.GetKernelStatus('total_num_virtual_procs')
85
86   # Set PRNG seed values:
87   if msd == None:
88     msd = rnd.randint(1000000)
89
```

```

90     msdrange = range(msd, msd + self.n_vp)
91     nest.SetKernelStatus({'rng_seed': msd + self.n_vp,
92                           'rng_seeds': msdrange})
93
94     def _build(self):
95         '''Create all nodes.'''
96
97         self.source_nodes = nest.Create('iaf_neuron', self.N_s)
98         self.target_nodes = nest.Create('iaf_neuron', self.N_t)
99
100    def _connect(self):
101        '''Connect all nodes.'''
102
103        nest.RandomDivergentConnect(self.source_nodes, self.target_nodes,
104                                     self.C, options={'allow_multapses': True})
105
106    def _get_degrees(self, msd):
107        '''
108            Reset NEST, create nodes, connect them, and retrieve the
109            resulting connections.
110
111            Parameters
112            -----
113            msd: Master RNG seed.
114
115            Return values
116            -----
117            list containing the in-degrees of the target nodes.
118
119            self._reset(msd)
120            self._build()
121            self._connect()
122
123        connections = nest.GetConnections(target=self.target_nodes)
124        target_connections = [conn[1] for conn in connections]
125        degrees = self._counter(target_connections)
126
127        return degrees
128
129    def _get_degrees_control(self, msd):
130        '''
131            Instead of using NEST, this method returns data with the expected
132            multinomial distribution.
133
134            Parameters
135            -----
136            msd: Master RNG seed.
137
138            Return values
139            -----
140            list containing the in-degrees of the target nodes.
141
142        ,,
143        if msd != None:
144            rnd.seed(msd)
145
146            con = rnd.randint(0, self.N_t, self.N_s * self.C)
147
148            degrees = self._counter(con)
149
150            return degrees
151
152    def chi_squared_test(self, msd=None, control=False):
153        '''
154            Create a single network and compare the resulting in-degree
155            distribution with the expected distribution using Pearson's chi-squared
156            GOF test.
157
158            Parameters
159            -----
160            msd : Master RNG seed.
161            control: Boolean value. If True, _get_degrees_control will
162            be used instead of _get_degrees.
163
164            Return values
165            -----
166            chi-squared statistic.
167            p-value from chi-squared test.
168
169            if control:
170                degrees = self._get_degrees_control(msd)
171            else:
172                degrees = self._get_degrees(msd)
173
174
175            # ddof: adjustment to the degrees of freedom. df = k-1-ddof
176            return scipy.stats.chisquare(numpy.array(degrees),
177                                         numpy.array(self.expected), ddof=0)
178
179    def two_level_test(self, n_runs, start_seed=None, control=False,
180                      show_histogram=False, histogram_bins=100,
181                      show_CDF=False):
182
183        '''
184            Create a network and run chi-squared GOF test n_runs times.
185            Test whether resulting p-values are uniformly distributed
186            on [0, 1] using the Kolmogorov-Smirnov GOF test.
187
188            Parameters
189            -----
190            n_runs : number of times to repeat chi-squared test.
191            start_seed : First PRNG seed value.
192            control : boolean value. If True, -get_degrees_control will be
193            used instead of -get_degrees.
194
195            show_histogram: Specify whether histogram should be displayed.
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
617
618
619
619
620
621
622
623
623
624
625
625
626
626
627
627
628
628
629
629
630
630
631
631
632
632
633
633
634
634
635
635
636
636
637
637
638
638
639
639
640
640
641
641
642
642
643
643
644
644
645
645
646
646
647
647
648
648
649
649
650
650
651
651
652
652
653
653
654
654
655
655
656
656
657
657
658
658
659
659
660
660
661
661
662
662
663
663
664
664
665
665
666
666
667
667
668
668
669
669
670
670
671
671
672
672
673
673
674
674
675
675
676
676
677
677
678
678
679
679
680
680
681
681
682
682
683
683
684
684
685
685
686
686
687
687
688
688
689
689
690
690
691
691
692
692
693
693
694
694
695
695
696
696
697
697
698
698
699
699
700
700
701
701
702
702
703
703
704
704
705
705
706
706
707
707
708
708
709
709
710
710
711
711
712
712
713
713
714
714
715
715
716
716
717
717
718
718
719
719
720
720
721
721
722
722
723
723
724
724
725
725
726
726
727
727
728
728
729
729
730
730
731
731
732
732
733
733
734
734
735
735
736
736
737
737
738
738
739
739
740
740
741
741
742
742
743
743
744
744
745
745
746
746
747
747
748
748
749
749
750
750
751
751
752
752
753
753
754
754
755
755
756
756
757
757
758
758
759
759
760
760
761
761
762
762
763
763
764
764
765
765
766
766
767
767
768
768
769
769
770
770
771
771
772
772
773
773
774
774
775
775
776
776
777
777
778
778
779
779
780
780
781
781
782
782
783
783
784
784
785
785
786
786
787
787
788
788
789
789
790
790
791
791
792
792
793
793
794
794
795
795
796
796
797
797
798
798
799
799
800
800
801
801
802
802
803
803
804
804
805
805
806
806
807
807
808
808
809
809
810
810
811
811
812
812
813
813
814
814
815
815
816
816
817
817
818
818
819
819
820
820
821
821
822
822
823
823
824
824
825
825
826
826
827
827
828
828
829
829
830
830
831
831
832
832
833
833
834
834
835
835
836
836
837
837
838
838
839
839
840
840
841
841
842
842
843
843
844
844
845
845
846
846
847
847
848
848
849
849
850
850
851
851
852
852
853
853
854
854
855
855
856
856
857
857
858
858
859
859
860
860
861
861
862
862
863
863
864
864
865
865
866
866
867
867
868
868
869
869
870
870
871
871
872
872
873
873
874
874
875
875
876
876
877
877
878
878
879
879
880
880
881
881
882
882
883
883
884
884
885
885
886
886
887
887
888
888
889
889
890
890
891
891
892
892
893
893
894
894
895
895
896
896
897
897
898
898
899
899
900
900
901
901
902
902
903
903
904
904
905
905
906
906
907
907
908
908
909
909
910
910
911
911
912
912
913
913
914
914
915
915
916
916
917
917
918
918
919
919
920
920
921
921
922
922
923
923
924
924
925
925
926
926
927
927
928
928
929
929
930
930
931
931
932
932
933
933
934
934
935
935
936
936
937
937
938
938
939
939
940
940
941
941
942
942
943
943
944
944
945
945
946
946
947
947
948
948
949
949
950
950
951
951
952
952
953
953
954
954
955
955
956
956
957
957
958
958
959
959
960
960
961
961
962
962
963
963
964
964
965
965
966
966
967
967
968
968
969
969
970
970
971
971
972
972
973
973
974
974
975
975
976
976
977
977
978
978
979
979
980
980
981
981
982
982
983
983
984
984
985
985
986
986
987
987
988
988
989
989
990
990
991
991
992
992
993
993
994
994
995
995
996
996
997
997
998
998
999
999
1000
1000
1001
1001
1002
1002
1003
1003
1004
1004
1005
1005
1006
1006
1007
1007
1008
1008
1009
1009
1010
1010
1011
1011
1012
1012
1013
1013
1014
1014
1015
1015
1016
1016
1017
1017
1018
1018
1019
1019
1020
1020
1021
1021
1022
1022
1023
1023
1024
1024
1025
1025
1026
1026
1027
1027
1028
1028
1029
1029
1030
1030
1031
1031
1032
1032
1033
1033
1034
1034
1035
1035
1036
1036
1037
1037
1038
1038
1039
1039
1040
1040
1041
1041
1042
1042
1043
1043
1044
1044
1045
1045
1046
1046
1047
1047
1048
1048
1049
1049
1050
1050
1051
1051
1052
1052
1053
1053
1054
1054
1055
1055
1056
1056
1057
1057
1058
1058
1059
1059
1060
1060
1061
1061
1062
1062
1063
1063
1064
1064
1065
1065
1066
1066
1067
1067
1068
1068
1069
1069
1070
1070
1071
1071
1072
1072
1073
1073
1074
1074
1075
1075
1076
1076
1077
1077
1078
1078
1079
1079
1080
1080
1081
1081
1082
1082
1083
1083
1084
1084
1085
1085
1086
1086
1087
1087
1088
1088
1089
1089
1090
1090
1091
1091
1092
1092
1093
1093
1094
1094
1095
1095
1096
1096
1097
1097
1098
1098
1099
1099
1100
1100
1101
1101
1102
1102
1103
1103
1104
1104
1105
1105
1106
1106
1107
1107
1108
1108
1109
1109
1110
1110
1111
1111
1112
1112
1113
1113
1114
1114
1115
1115
1116
1116
1117
1117
1118
1118
1119
1119
1120
1120
1121
1121
1122
1122
1123
1123
1124
1124
1125
1125
1126
1126
1127
1127
1128
1128
1129
1129
1130
1130
1131
1131
1132
1132
1133
1133
1134
1134
1135
1135
1136
1136
1137
1137
1138
1138
1139
1139
1140
1140
1141
1141
1142
1142
1143
1143
1144
1144
1145
1145
1146
1146
1147
1147
1148
1148
1149
1149
1150
1150
1151
1151
1152
1152
1153
1153
1154
1154
1155
1155
1156
1156
1157
1157
1158
1158
1159
1159
1160
1160
1161
1161
1162
1162
1163
1163
1164
1164
1165
1165
1166
1166
1167
1167
1168
1168
1169
1169
1170
1170
1171
1171
1172
1172
1173
1173
1174
1174
1175
1175
1176
1176
1177
1177
1178
1178
1179
1179
1180
1180
1181
1181
1182
1182
1183
1183
1184
1184
1185
1185
1186
1186
1187
1187
1188
1188
1189
1189
1190
1190
1191
1191
1192
1192
1193
1193
1194
1194
1195
1195
1196
1196
1197
1197
1198
1198
1199
1199
1200
1200
1201
1201
1202
1202
1203
1203
1204
1204
1205
1205
1206
1206
1207
1207
1208
1208
1209
1209
1210
1210
1211
1211
1212
1212
1213
1213
1214
1214
1215
1215
1216
1216
1217
1217
1218
1218
1219
1219
1220
1220
1221
1221
1222
1222
1223
1223
1224
1224
1225
1225
1226
1226
1227
1227
1228
1228
1229
1229
1230
1230
1231
1231
1232
1232
1233
1233
1234
1234
1235
1235
1236
1236
1237
1237
1238
1238
1239
1239
1240
1240
1241
1241
1242
1242
1243
1243
1244
1244
1245
1245
1246
1246
1247
1247
1248
1248
1249
1249
1250
1250
1251
1251
1252
1252
1253
1253
1254
1254
1255
1255
1256
1256
1257
1257
1258
1258
1259
1259
1260
1260
1261
1261
1262
1262
1263
1263
1264
1264
1265
1265
1266
1266
1267
1267
1268
1268
1269
1269
1270
1270
1271
1271
1272
1272
1273
1273
1274
1274
1275
1275
1276
1276
1277
1277
1278
1278
1279
1279
1280
1280
1281
1281
1282
1282
1283
1283
1284
1284
1285
1285
1286
1286
1287
1287
1288
1288
1289
1289
1290
1290
1291
1291
1292
1292
1293
1293
1294
1294
1295
1295
1296
1296
1297
1297
1298
1298
1299
1299
1300
1300
1301
1301
1302
1302
1303
1303
1304
1304
1305
1305
1306
1306
1307
1307
1308
1308
1309
1309
1310
1310
1311
1311
1312
1312
1313
1313
1314
1314
1315
1315
1316
1316
1317
1317
1318
1318
1319
1319
1320
1320
1321
1321
1322
1322
1323
1323
1324
1324
1325
1325
1326
1326
1327
1327
1328
1328
1329
1329
1330
1330
1331
1331
1332
1332
1333
1333
1334
1334
1335
1335
1336
1336
1337
1337
1338
1338
1339
1339
1340
1340
1341
1341
1342
1342
1343
1343
1344
1344
1345
1345
1346
1346
1347
1347
1348
1348
1349
1349
1350
1350
1351
1351
1352
1352
1353
1353
1354
1354
1355
1355
1356
1356
1357
1357
1358
1358
1359
1359
1360
1360
1361
1361
1362
1362
1363
1363
1364
1364
1365
1365
1366
1366
1367
1367
1368
1368
1369
1369
1370
1370
1371
1371
1372
1372
1373
1373
1374
1374
1375
1375
1376
1376
1377
1377
1378
1378
1379
1379
1380
1380
1381
1381
1382
1382
1383
1383
1384
1384
1385
1385
1386
1386
1387
1387
1388
1388
1389
1389
1390
1390
1391
1391
1392
1392
1393
1393
1394
1394
1395
1395
1396
1396
1397
1397
1398
1398
1399
1399
1400
1400
1401
1401
1402
1402
1403
1403
1404
1404
1405
1405
1406
1406
1407
1407
1408
1408
1409
1409
1410
1410
1411
1411
1412
1412
1413
1413
1414
1414
1415
1415
1416
1416
1417
1417
1418
1418
1419
1419
1420
1420
1421
1421
1422
1422
1423
1423
1424
1424
1425
1425
1426
1426
1427
1427
1428
1428
1429
1429
1430
1430
1431
1431
1432
1432
1433
1433
1434
1434
1435
1435
1436
1436
1437
1437
1438
1438
1439
1439
1440
1440
1441
1441
1442
1442
1443
1443
1444
1444
1445
1445
1446
1446
1447
1447
1448
1448
1449
1449
1450
1450
1451
1451
1452
1452
1453
1453
1454
1454
1455
1455
1456
1456
1457
1457
1458
1458
1459
1459
1460
1460
1461
1461
1462
1462
1463
1463
1464
1464
1465
1465
1466
1466
1467
1467
1468
1468
1469
1469
1470
1470
1471
1471
1472
1472
1473
1473
1474
1474
1475
1475
1476
1476
1477
1477
1478
1478
1479
1479
1480
1480
1481
1481
1482
1482
1483
1483
1484
1484
1485
1485
1486
1486
1487
1487
1488
1488
1489
1489
1490
1490
1491
1491
1492
1492
1493
1493
1494
1494
1495
1495
1496
1496
1497
1497
1498
1498
1499
1499
1500
1500
1501
1501
1502
1502
1503
1503
1504
1504
1505
1505
1506
1506
1507
1507
1508
1508
1509
1509
1510
1510
1511
1511
1512
1512
1513
1513
1514
1514
1515
1515
1516
1516
1517
1517
1518
1518
1519
1519
1520
1520
1521
1521
1522
1522
1523
1523
1524
1524
1525
1525
1526
1526
1527
1527
1528
1528
1529
1529
1530
1530
1531
1531
1532
15
```

```

194     histogram_bins: Number of histogram bins.
195     show_CDF : Specify whether EDF should be displayed.
196
197     Return values
198
199     -----
200     KS statistic.
201     p-value from KS test.
202
203     self.pvalues = []
204
205     if start_seed == None:
206         for i in range(n_runs):
207             print 'Running test %d, %d, %d, %d' % (i + 1, n_runs)
208             chi, p = self.chi_squared_test(None, control)
209             self.pvalues.append(p)
210
211     else:
212         seed_jump = self.n_vp + 1
213         end_seed = start_seed + n_runs * seed_jump
214         for seed in range(start_seed, end_seed, seed_jump):
215             print 'Running test %d, %d, %d, %d' % \
216                 (1 + (seed - start_seed) / seed_jump, n_runs)
217             chi, p = self.chi_squared_test(seed, control)
218             self.pvalues.append(p)
219
220     ks, p = scipy.stats.ks_2samp(self.pvalues, 'uniform',
221                                 alternative='two-sided')
222
223     if show_CDF:
224         plt.figure()
225         self.pvalues.sort()
226         y = [(i + 1.) / len(self.pvalues)
227               for i in range(len(self.pvalues))]
228         plt.step([0.0] + self.pvalues + [1.0], [0.0] + y + [1.0])
229         plt.xlabel('P-values')
230         plt.ylabel('Empirical distribution function')
231
232     if show_histogram:
233         plt.figure()
234         plt.hist(self.pvalues, bins=histogram_bins)
235         plt.xlabel('P-values')
236         plt.ylabel('Frequency')
237
238     if show_CDF or show_histogram:
239         plt.show(block=True)
240
241     return ks, p
242
243     if __name__ == '__main__':
244         test = RDC_tester(N_s=1000, N_t=1000, C=1000)
245         ks, p = test.two_level_test(n_runs=1000, start_seed=0)

```



## Appendix D

## Automated test suite for random convergent and divergent connections

```

1 ,,
2 @author: Daniel Hjortholm
3
4 Unitests for RandomConvergentConnect and RandomDivergentConnect .
5 ,,
6 import unittest
7
8 from test_RCC import RCC_tester
9 from test_RDC import RDC_tester
10
11
12 class RCDCTestCase(unittest.TestCase):
13     ,,'Statistical tests for Random{Con,Di}vergentConnect.''
14
15     def setUp(self):
16         ,,'Set test parameters and critical values . ,,
17
18         # Small network
19         self.N_small = 10 # Number of nodes
20         self.C_small = 100 # Average number of connections per node
21         self.n_small = 1000 # Number of times to repeat test
22
23         # Medium sized network
24         self.N_medium = 100 # Number of nodes
25         self.C_medium = 100 # Average number of connections per node
26         self.n_medium = 100 # Number of times to repeat test
27
28         # Large network
29         self.N_large = 1000 # Number of nodes
30         self.C_large = 1000 # Average number of connections per node
31         self.n_large = 100 # Number of times to repeat test
32
33         # Critical values
34         self.alpha1_lower = 0.025
35         self.alpha1_upper = 0.975
36         self.alpha2 = 0.05
37
38     def adaptive_test(self, test, n_runs):
39
40         ,,'Create a single network using Random{Con,Di}vergentConnect
41
42 and run a chi-squared GOF test on the connection distribution.
43 If the result is extreme (high or low), run a two-level test.
44
45 Parameters
46 -----
47     test : Instance of RCC_tester or RDC_tester class.
48     n_runs: If chi-square test fails, test is repeated n_runs times,
49     and the KS test is used to analyze results.
50
51     Return values
52     -----
53     boolean value. True if test was passed, False otherwise.
54     ,
55     chi, p = test.chi_squared_test(msd=None)
56
57     if self.alpha1_lower < p < self.alpha1_upper:
58
59         return True
60
61     ks, p = test.two_level_test(n_runs=n_runs, start_seed=None)
62
63     return True if p > self.alpha2 else False
64
65     ,,'Statistical test of RandomConvergentConnect with a small network'':
66
67     test = RCC_tester(N_s=self.N_small, N_t=self.N_small, C=self.C_small)
68     passed = self.adaptive_test(test, n_runs=self.n_small)
69     self.assertTrue(passed, 'RandomConvergentConnect did not , \
70         ,pass the statistical test procedure.')
71
72     def test_RCC_large(self):
73
74         ,,'Statistical test of RandomConvergentConnect with a large network'':
75
76         test = RCC_tester(N_s=self.N_large, N_t=self.N_large, C=self.C_large)
77         passed = self.adaptive_test(test, n_runs=self.n_large)
78         self.assertTrue(passed, 'RandomConvergentConnect did not , \
79             ,pass the statistical test procedure.')
80
81     ,,'Statistical test of RandomConvergentConnect with 4 threads'':
82
83

```

```

83     test = RCC_tester(N_s=self.N_medium, N_t=self.N_medium, C=self.C_medium,
84                         threads=4)
85     passed = self.adaptive_test(test, n_runs=self.n_medium)
86     self.assertTrue(passed, 'RandomConvergentConnect did not , \
87                     \n'pass the statistical test procedure.')
88
89     def test_RDC_small(self):
90         '''Statistical test of RandomDivergentConnect with a small network'''
91
92         test = RDC_tester(N_s=self.N_small, N_t=self.N_small, C=self.C_small)
93         passed = self.adaptive_test(test, n_runs=self.n_small)
94         self.assertTrue(passed, 'RandomDivergentConnect did not , \
95                     \n'pass the statistical test procedure.')
96
97     def test_RDC_large(self):
98         '''Statistical test of RandomDivergentConnect with a large network'''
99
100        test = RDC_tester(N_s=self.N_large, N_t=self.N_large, C=self.C_large)
101        passed = self.adaptive_test(test, n_runs=self.n_large)
102        self.assertTrue(passed, 'RandomDivergentConnect did not , \
103                     \n'pass the statistical test procedure.')
104
105    def test_RDC_threaded(self):
106        '''Statistical test of RandomDivergentConnect with 4 threads'''
107
108        test = RDC_tester(N_s=self.N_medium, N_t=self.N_medium, C=self.C_medium,
109                         threads=4)
110        passed = self.adaptive_test(test, n_runs=self.n_medium)
111        self.assertTrue(passed, 'RandomDivergentConnect did not , \
112                     \n'pass the statistical test procedure.')
113
114
115    def suite():
116        suite = unittest.makeSuite(RCDCTestCase, 'test')
117        return suite
118
119    if __name__ == '__main__':
120        runner = unittest.TextTestRunner(verbosity=2)
121        runner.run(suite())
122

```

# Appendix E

## Test script for 2D spatially structured network

```

1   ,
2   @author: Daniel Hjertholm
3
4 Tests for ConnectLayers .
5   ,
6
7 import numpy as np
8 import numpy.random as rnd
9 import scipy.integrate
10 import scipy.stats
11 import matplotlib.pyplot as plt
12 import nest
13 import nest.topology as topo
14
15
16 class ConnectLayers2D_tester(object):
17     '''Class used for testing ConnectLayers .'''
18
19     def __init__(self, L, N, kernel_name, kernel_params=None,
20                  source_pos=None, msd=None, threads=1):
21         '''
22             Initialize an ConnectLayers2D_tester object.
23
24 Sets up the experiment, and defines network layer and
25 connection parameters .
26
27 Parameters
28 -----
29     L : Side length of square layer.
30     N : Number of nodes.
31     kernel_name : Name of distance dependent probability
32                  function (kernel) to test.
33     kernel_params: Parameters for kernel function. If
34                  omitted, sensible defaults are calculated
35                  based on layer size.
36     source_pos : Source node position. Default is center.
37     msd : Master PRNG seed. Default is None.
38     threads : Number of local threads. Default is 1.
39
40
41     self.threads = threads
42     self.msd = msd
43     self.L = float(L)
44     self.N = N
45
46         self.kernel_name = kernel_name
47         kernels = {
48             'constant': self._constant,
49             'linear': self._linear,
50             'exponential': self._exponential,
51             'gaussian': self._gauss}
52
53         default_params = {
54             'constant': 1.0,
55             'linear': {'a': np.sqrt(2) / self.L, 'c': 1.0},
56             'exponential': {'a': 1.0, 'c': 0.0,
57                             'tau': -self.L /
58                             (np.sqrt(2) * np.log((1 - 0) / 1))},
59             'gaussian': {'p_center': 1., 'sigma': self.L / 4.,
60                          'mean': 0., 'c': 0.}}
61
62         self.params = default_params[kernel_name]
63
64         if kernel_params is not None:
65             if kernel_name == 'constant':
66                 self.params = kernel_params
67             else:
68                 self.params.update(kernel_params)
69
70         self.kernel_func = kernels[kernel_name]
71
72         if source_pos is None: source_pos = (0., 0.)
73         self.ldict_s = {'elements': 'iaf_neuron', 'positions': [source_pos],
74                         'extent': [self.L, self.L], 'edge_wrap': True}
75
76         if msd is not None:
77             rnd.seed(msd)
78             x = rnd.uniform(-self.L / 2., self.L / 2., self.N)
79             y = rnd.uniform(-self.L / 2., self.L / 2., self.N)
80             pos = zip(x, y)
81             self.ldict_t = {'elements': 'iaf_neuron', 'positions': pos,
82                           'extent': [self.L, self.L], 'edge_wrap': True}
83
84         if kernel_name == 'constant':
85             self.kernel = self.params
86
87         else:
88             self.kernel = {'self.kernel_name': self.params}
89             self.conndict = {'connection_type': 'divergent',

```

```

90   'mask': self.mask, 'kernel': self.kernel}.
91
92   def _constant(self, D):
93     ,,'Constant kernel function.,,
94
95     return self.params
96
97   def _linear(self, D):
98     ,,'Linear kernel function.,,
99
100    return self.params['c'] + self.params['a'] * D
101
102   def _exponential(self, D):
103     ,,'Exponential kernel function.,,
104
105    return (self.params['c'] + self.params['a'] *
106      np.e ** (-D / self.params['tau'])))
107
108   def _gauss(self, D):
109     ,,'Gaussian kernel function.,,
110
111    return (self.params['c'] + self.params['p_center'] *
112      np.e ** -( (D - self.params['mean']) ** 2 /
113        (2. * self.params['sigma'] ** 2)))
114
115   def _pdf(self, D):
116     ,,
117
118     Unnormalized probability density function (PDF).
119
120   Parameters
121
122   -----
123   D: Distance in interval [0, L/sqrt(2)].
124
125   -----
126   Return values
127
128   -----
129   if D <= self.L / 2.:
130     return (max(0., min(1., self.kernel_func(D)) * np.pi * D)
131       elif self.L / 2. < D <= self.L / np.sqrt(2):
132         return (max(0., min(1., self.kernel_func(D)) *
133           D * (np.pi - 4. * np.arccos(self.L / (D * 2.))))
134
135   else:
136     return 0.
137
138   def _cdf(self, D):
139
140   Normalized cumulative distribution function (CDF).
141
142   D: Iterable of distances in interval [0, L/sqrt(2)].
143   Returns
144   -----
145   ,,, List of CDF(d) for each distance d in D.
146
147   ,,
148
149   cdf = []
150   last_d = 0.
151   for d in D:
152     cdf.append(scipy.integrate.quad(self._pdf, last_d, d)[0])
153   last_d = d
154
155   cdf = np.cumsum(cdf)
156
157   top = scipy.integrate.quad(self._pdf, 0, self.L / np.sqrt(2))[0]
158   normed_cdf = cdf / top
159
160   # Stored in case the CDF is to be plotted later.
161
162   return normed_cdf
163
164   def _reset(self):
165     ,,'Reset the NEST kernel and set PRNG seed values.,,
166
167   nest.ResetKernel()
168   nest.SetKernelStatus({'local_num_threads': self.threads})
169   self.n_vp = nest.GetKernelStatus('total_num_virtual_procs')
170
171   # Set PRNG seed values:
172   if self.msd is None:
173     msd = rnd.randint(1000000)
174
175   else:
176     msd = self.msd + 1 # The first was used in __init__.
177   msdrange = range(msd, msd + self.n_vp)
178   nest.SetKernelStatus({'rng_seed': msd + self.n_vp,
179     'rng_seeds': msdrange})
180
181   def _build(self):
182     ,,'Create layers.,,
183
184   self.ls = topo.CreateLayer(self.lsdict_s)
185   self.lt = topo.CreateLayer(self.ltdict_t)
186   self.driver = topo.FindCenterElement(self.ls)
187
188   def _connect(self):
189     ,,'Connect layers.,,
190
191   topo.ConnectLayers(self.ls, self.lt, self.conndict)
192
193   def _get_distances(self):
194
195   -----

```

```

194     ,,
195     Create and connect layers, and get distances to connected nodes.
196
197     Return values
198     -----
199     Ordered list of distances to connected nodes.
200
201     self._reset()
202     self._buid()
203     self._connect()
204
205     if control:
206         connections = nest.GetConnections(source=nest.driver)
207         target_nodes = [conn[1] for conn in connections]
208         if len(target_nodes) == 0: return None, None
209         dist = topo.Distance(self.driver, target_nodes)
210         dist.sort()
211
212     return dist
213
214     def _get_distances_control(self):
215         ,,
216         Instead of using NEST's actual connection algorithm, this
217         method creates data with the expected distribution.
218
219     Return values
220     -----
221     Ordered list of distances to connected nodes.
222
223     self._reset()
224     self._buid()
225
226     dist = topo.Distance(self.driver, nest.GetLeaves(self.lt)[0])
227     dist = [d for d in dist if rnd.uniform() < self.kernel_func(d)]
228     dist.sort()
229
230     return dist
231
232     def ks_test(self, control=False, alternative='two_sided',
233                 show_network=False, show_CDF=False, show_PDF=False,
234                 histogram_bins=100):
235
236         ,,
237         Perform a Kolmogorov-Smirnov GOF test on the distribution
238         of distances to connected nodes.
239
240         Parameters
241         -----
242             control : Boolean value. If True, -get_distances-control
243             will be used instead of -get-distances.
244             alternative : Alternative hypothesis. 'two-sided' (default),
245             'less' or 'greater'.
246
247             show_network : Specify whether network plot should be displayed.
248             show_CDF : Specify whether CDF should be displayed.
249             show_PDF : Specify whether PDF should be displayed.
250             histogram_bins: Number of histogram bins for PDF plot.
251
252     Return values
253     -----
254         KS statistic.
255         p-value from KS test.
256
257         if control:
258             dist = self._get_distances_control()
259
260         dist = self._get_distances()
261
262         ks, p = scipy.stats.ks-test(dist, self._cdf,
263                                     alternative=alternative)
264
265         if show_network and not control:
266             # Adjust size of nodes in plot based on number of nodes.
267             nodesize = max(1, int(round(11. / 11 - self.N / 11000.)))
268
269         fig = topo.PlotLayer(self.lt, nodesize=nodesize,
270                             nodecolor='grey')
271
272         # Only the gaussian kernel can be plotted.
273         if self.kernel_name == 'gaussian':
274             topo.PlotTargets(self.driver, self.lt, fig=fig,
275                             mask=self.mask, kernel=self.kernel,
276                             mask_color='purple',
277                             kernel_color='blue',
278                             src_size=50, src_color='black',
279                             tgt_size=nodesize, tgt_color='red')
280
281         topo.PlotTargets(self.driver, self.lt, fig=fig,
282                         mask=self.mask, mask_color='purple',
283                         src_size=50, src_color='black',
284                         tgt_size=nodesize, tgt_color='red')
285
286         if show_CDF:
287             plt.figure()
288             plt.plot(dist, self.cdf_list, ',', color='black', linewidth=3,
289             label='Theory', zorder=1)
290             y = [(i + 1.) / len(dist) for i in range(len(dist))]
291             plt.step([0.0] + dist, [0.0] + y, color='red',
292                     linewidth=1, label='Empirical', zorder=2)
293             plt.ylim(0, 1)
294
295             plt.xlabel('Distance')
296             plt.ylabel('CDF')
297             plt.legend(loc='center right')

```

```

298 if show_PDF:
299     plt.figure()
300     x = np.linspace(0, self.L / np.sqrt(2), 1000)
301     area = scipy.integrate.quad(self._pdf, 0, self.L / np.sqrt(2)) [0]
302     y = np.array([self._pdf(D) for D in x]) / area
303     plt.plot(x, y, color='black', linewidth=3, label='Theory', zorder=1)
304     plt.hist(dist, bins=nistogram_bins, histtype='step',
305             linewidth=1, normed=True, color='red',
306             label='Empirical', zorder=2)
307     plt.ylim(ymin=0.)
308
309     plt.xlabel('Distance')
310     plt.ylabel('PDF')
311     plt.legend(loc='center right')
312
313     if show_CDF or show_PDF or show_network:
314         plt.show(block=True)
315
316     return ks
317
318 def z_test(self, control=False):
319     """
320     Perform a Z-test on the total number of connections.
321
322     Parameters
323
324     -----
325     control: Boolean value. If True, -get_distances_control will be used
326     instead of -get_distances.
327
328     Returns
329     -----
330     Standard score (z-score).
331     Two-sided p-value.
332     ,
333
334     if control:
335         num = len(self._get_distances_control())
336     else:
337         num = len(self._get_distances())
338
339     dist = topo.Distance(self.driver, nest.GetLeaves(self.lt)[0])
340     ps = ([max(0., min(1., self.kernel_func(D))) for D in dist])
341     exp = sum(ps)
342     var = sum([p * (1. - p) for p in ps])
343     if var == 0: return np.nan, 1.0
344     sd = np.sqrt(var)
345     z = abs((num - exp) / sd)
346     p = 2. * (1. - scipy.stats.norm.cdf(z))
347
348     return z, p
349

```

# Appendix F

## Test script for 3D spatially structured network

```

1   ,
2   @author: Daniel Hjertholm
3
4 Tests for ConnectLayers .
5   ,
6
7 import numpy as np
8 import numpy.random as rnd
9 import scipy.integrate
10 import scipy.stats
11 import matplotlib.pyplot as plt
12 import nest
13 import nest.topology as topo
14
15
16 class ConnectLayers3D_tester(object):
17     """Class used for testing ConnectLayers ."""
18
19     def __init__(self, L, N, kernel_name, kernel_params=None,
20                  source_pos=None, msd=None, threads=1):
21         """
22             Initialize an ConnectLayers3D_tester object.
23
24 Sets up the experiment, and defines network layer and
25 connection parameters .
26
27 Parameters
28 -----
29     L : Side length of cubic layer.
30     N : Number of nodes.
31     kernel_name : Name of distance dependent probability
32                 function (kernel) to test.
33     kernel_params : Parameters for kernel function. If
34                 omitted, sensible defaults are calculated
35                 based on layer size.
36     source_pos : Source node position. Default is center.
37     msd : Master PRNG seed. Default is None.
38     threads : Number of local threads. Default is 1.
39
40
41     self.threads = threads
42     self.msd = msd
43     self.L = float(L)
44     self.N = N
45
46         self.kernel_name = kernel_name
47         kernels = {
48             'constant': self._constant,
49             'linear': self._linear,
50             'exponential': self._exponential,
51             'gaussian': self._gauss}
52
53         default_params = {
54             'constant': 1.0,
55             'linear': {'a': -2.0 / (np.sqrt(3) * self.L), 'c': 1.0},
56             'exponential': {'a': 1.0, 'c': 0.0,
57                             'tau': -self.L * np.sqrt(3) /
58                             (2.0 * np.log((.1 - 0) / 1))},
59             'gaussian': {'p_center': 1., 'sigma': self.L / 4.,
60                          'mean': 0., 'c': 0.}}
61
62         self.params = default_params[kernel_name]
63
64         if kernel_params is not None:
65             if kernel_name == 'constant':
66                 self.params = kernel_params
67             else:
68                 self.params.update(kernel_params)
69
70         self.kernel_func = kernels[kernel_name]
71
72         if source_pos is None: source_pos = (0., 0., 0.)
73         self.ldict_s = {'elements': 'iaf_neuron', 'positions': [source_pos],
74                         'extent': [self.L, self.L, self.L], 'edge_wrap': True}
75
76         if msd is not None:
77             rnd.seed(msd)
78             x = rnd.uniform(-self.L / 2., self.L / 2., self.N)
79             y = rnd.uniform(-self.L / 2., self.L / 2., self.N)
80             z = rnd.uniform(-self.L / 2., self.L / 2., self.N)
81             pos = zip(x, y, z)
82             self.ldict_t = {'elements': 'iaf_neuron', 'positions': pos,
83                           'extent': [self.L, self.L, self.L], 'edge_wrap': True}
84
85         self.mask = {'box': {'lower_left': [-self.L / 2.,
86                               -self.L / 2., -self.L / 2.],
87                               'upper_right': [self.L / 2.,
88                               self.L / 2., self.L / 2.]}}
89
90
91         if kernel_name == 'constant':
92             self.kernel = self.params
93
94

```



```

194
def _build(self):
195     , , , Create layers , ,
196
197     self.ls = topo.CreateLayer(self.Idict_s)
198     self.lt = topo.CreateLayer(self.Idict_t)
199     self.driver = topo.FindCenterElement(self.ls)
200
201     def connect(self):
202         , , , Connect layers , ,
203
204         topo.ConnectLayers(self.ls, self.lt, self.comndict)
205
206     def _get_distances(self):
207         , ,
208
209         Create and connect layers, and get distances to connected nodes.
210
211         Return values
212
213         Ordered list of distances to connected nodes.
214
215         self._reset()
216         self._build()
217         self._connect()
218
219         connections = nest.GetConnections(source=self.driver)
220         target_nodes = [conn[1] for conn in connections]
221         if len(target_nodes) == 0: return None, None
222         dist = topo.Distance(self.driver, target_nodes)
223         dist.sort()
224
225         return dist
226
227     def _get_distances_control(self):
228         , ,
229
230         Instead of using NEST's actual connection algorithm, this
231         method creates data with the expected distribution.
232
233         Return values
234
235         Ordered list of distances to connected nodes.
236
237         self._reset()
238         self._build()
239
240         dist = topo.Distance(self.driver, nest.GetLeaves(self.lt)[0])
241         dist = [d for d in dist if rnd.uniform() < self.kernel_func(d)]
242         dist.sort()
243
244         return dist
245
246     def ks_test(self, control=False, alternative='two-sided',
247     show_network=False, show_CDF=False, show_PDF=False,
248     histogram_bins=100):
249
250         , ,
251
252         Perform a Kolmogorov-Smirnov GOF test on the distribution
253         of distances to connected nodes.
254
255         Parameters
256         -----
257             control : Boolean value. If True, -get_distances_control
258             will be used instead of -get_distances .
259             alternative : Alternative hypothesis. 'two_sided' (default),
260             'less' or 'greater'.
261             show_network : Specify whether network plot should be displayed.
262             show_CDF : Specify whether CDF should be displayed.
263             show_PDF : Specify whether PDF should be displayed.
264             histogram_bins: Number of histogram bins for PDF plot.
265
266         Return values
267
268             KS statistic.
269             p-value from KS test.
270
271         if control:
272             dist = self._get_distances_control()
273
274         else:
275             dist = self._get_distances()
276
277         ks, p = scipy.stats.kstest(dist, self._cdf,
278             alternative='grey')
279
280         if show_network and not control:
281             # Adjust size of nodes in plot based on number of nodes
282             nodesize = max(1, int(round(111. / 11 - self.N / 11000.)))
283
284         fig = topo.PlotLayer(self.lt, nodesize=nodesize,
285             nodecolor='grey')
286
287         # Only the gaussian kernel can be plotted.
288         if self.kernel_name == 'gaussian':
289             topo.PlotTargets(self.driver, self.lt, fig=fig,
290                 mask=self.mask, kernel=self.kernel,
291                 mask_color='purple',
292                 kernel_color='blue',
293                 src_size=50, src_color='black',
294                 tgt_size=nodesize, tgt_color='red')
295
296         topo.PlotTargets(self.driver, self.lt, fig=fig,
297             mask=self.mask, mask_color='purple',
298             src_size=50, src_color='black',
299             src_size=50, src_color='black',
300             src_size=50, src_color='black')

```

```

298     tgt_size=nodesize, tgt_color='red')
299
300     if show_CDF():
301         plt.figure()
302             plt.plot(dist, self.cdf_list, ' ', color='black', linewidth=3,
303                         label='Theory', zorder=1)
304                 y = [(i + 1.) / len(dist) for i in range(len(dist))]
305                     plt.step([0.0] + dist, [0.0] + y, color='red',
306                                 linewidth=1, label='Empirical', zorder=2)
307
308                     plt.ylim(0, 1)
309
310                     plt.xlabel('Distance')
311                     plt.ylabel('CDF')
312
313                     plt.legend(loc='center right')
314
315             if show_PDF():
316                 plt.figure()
317                     x = np.linspace(0, self.L * np.sqrt(3) / 2, 1000)
318                         area = scipy.integrate.quad(self.pdf, 0,
319                                         self.L * np.sqrt(3) / 2) [0]
320                             y = np.array([self.pdf(D) for D in x]) / area
321                             plt.plot(x, y, color='black', linewidth=3, label='Theory', zorder=1)
322                             plt.hist(dist, bins=histogram_bins, histtype='step',
323                                         linewidth=1, normed=True, color='red',
324                                         label='Empirical', zorder=2)
325
326             plt.ylim(ymin=0.)
327
328             plt.xlabel('Distance')
329             plt.ylabel('PDF')
330             plt.legend(loc='center right')
331
332             if show_CDF or show_PDF or show_network:
333
334             plt.show(block=True)
335
336             return ks, p
337
338             def z_test(self, control=False):
339                 ,,
340                     control: Boolean value. If True, -get_instances_control will be used
341                     instead of -get_instances.
342
343             Parameters
344             -----
345                 Standard score (z-score).
346                 Two-sided p-value.
347
348             if control:
349

```

# Appendix G

## Automated test suite for spatially structured networks

```
1 ,,
2 @author: Daniel Hjortholm
3
4 Unittests for 2D and 3D spatially structured networks.
5 ,,
6
7 import unittest
8 import scipy.stats
9
10 from test_2D import ConnectLayers2D_tester
11 from test_3D import ConnectLayers3D_tester
12
13
14 class ConnectLayersTestCase(unittest.TestCase):
15     """Statistical tests for ConnectLayers."""
16
17     def setUp(self):
18         """Set test parameters and critical values."""
19         self.N = 100000 # Number of nodes
20         self.L = 1.0 # Layer size.
21         self.n_runs = 100 # Number of times to repeat test
22
23         # Critical values
24         self.alpha1 = 0.01
25         self.alpha2 = 0.01
26
27         def ks_test(self, tester_class, kernel_name, kernel_params={} ,
28                     source_pos=None, threads=1):
29             """
30             Create a single network using ConnectLayers, and perform a Z-test on the
31             total connection count. If the result is suspicious, the test is
32             repeated n_runs times, and the resulting p-values are compared with the
33             expected uniform distribution using the KS test.
34
35             Parameters
36             tester_class : ConnectLayers2D-tester or ConnectLayers3D-tester.
37             kernel_name : Name of distance dependent probability
38             function (kernel) to test.
39             tester-class : ConnectLayers2D-tester or ConnectLayers3D-tester.
40             kernel_name : Name of distance dependent probability
41             function (kernel) to test.
42             kernel_params: Parameters for kernel function. Optional.
43             source_pos : Source node position. Optional. Default is center.
44             threads : Number of local threads. Optional. Default is 1.
45
46         Return values
47         -----
48         boolean value. True if test was passed, False otherwise.
49 ,,
50         test = tester_class(L=self.L, N=self.N, kernel_name=kernel_name,
51                             kernel_params=kernel_params,
52                             source_pos=source_pos, msd=None)
53
54         ks, p = test.ks_test()
55
56         if p > self.alpha1:
57             return True
58         else:
59             ps = []
60             print '',
61             for i in range(self.n_runs):
62                 print 'Running test %d, % (i + 1, self.n_runs)
63                 test = tester_class(L=self.L, N=self.N,
64                                     kernel_name=kernel_name,
65                                     kernel_params=kernel_params,
66                                     source_pos=source_pos, msd=None)
67
68                 ps.append(test.ks_test()[1])
69
70             ks, p = scipy.stats.ks_2samp(ps, 'uniform', alternative='two-sided')
71
72             return True if p > self.alpha2 else False
73
74         def z_test(self, tester_class, kernel_name, kernel_params={} ,
75                     source_pos=None, threads=1):
76             """
77             Create a single network using ConnectLayers, and perform a Z-test on the
78             total connection count. If the result is suspicious, the test is
79             repeated n_runs times, and the resulting p-values are compared with the
80             expected uniform distribution using the KS test.
81
82             Parameters
83             -----
84             tester-class : ConnectLayers2D-tester or ConnectLayers3D-tester.
85             kernel_params: Parameters for kernel function. Optional.
86             source_pos : Source node position. Optional. Default is center.
87             threads : Number of local threads. Optional. Default is 1.
88
89             Return values
90             -----
```

```

90     boolean value. True if test was passed, False otherwise.
91
92     test = tester_class(L=self.L, N=self.N, kernel_name=kernel_name,
93                         kernel_params=kernel_params,
94                         source_pos=source_pos, msd=None)
95
96     z, p = test.z_test()
97
98     if p > self.alpha1:
99         return True
100    else:
101        ps = []
102        print '',
103        for i in range(self.n_runs):
104            print 'Running test %d of %d.' % (i + 1, self.n_runs)
105            test = tester_class(L=self.L, N=self.N,
106                                kernel_name=kernel_name,
107                                kernel_params=kernel_params,
108                                source_pos=source_pos, msd=None)
109            ps.append(test.z_test() [1])
110            z, p = scipy.stats.kstest(ps, 'uniform', alternative='two-sided')
111            if p > self.alpha2 else False
112
113        def test_2D_constant_ks(self):
114            '''KS test performed on source-target node distances'''
115            self.assertTrue(self.ks_test(ConnectLayers2DTester, 'constant',
116                                         kernel_params=0.5),
117                                         'ConnectLayers failed to pass the KS test.')
118
119        def test_2D_constant_z(self):
120            '''Z-test performed on total connection count'''
121            self.assertTrue(self.z_test(ConnectLayers2DTester, 'constant',
122                                      kernel_params=0.5),
123                                      'ConnectLayers failed to pass the Z-test.')
124
125        def test_2D_linear_ks(self):
126            '''KS test performed on source-target node distances'''
127            self.assertTrue(self.ks_test(ConnectLayers2DTester, 'linear'),
128                                         'ConnectLayers failed to pass the KS test.')
129
130        def test_2D_linear_z(self):
131            '''Z-test performed on total connection count'''
132            self.assertTrue(self.z_test(ConnectLayers2DTester, 'linear'),
133                                         'ConnectLayers failed to pass the Z-test.')
134
135        def test_2D_exponential_ks(self):
136            '''KS test performed on source-target node distances'''
137
138        def test_3D_constant_ks(self):
139            '''KS test performed on source-target node distances'''
140
141
142        self.assertTrue(self.ks_test(ConnectLayers2DTester, 'exponential'),
143                      'ConnectLayers failed to pass the KS test.')
144
145        def test_2D_exponential_z(self):
146            '''Z-test performed on total connection count'''
147            self.assertTrue(self.z_test(ConnectLayers2DTester, 'exponential'),
148                          'ConnectLayers failed to pass the Z-test')
149
150        def test_2D_gaussian_ks(self):
151            '''KS test performed on source-target node distances'''
152            self.assertTrue(self.ks_test(ConnectLayers2DTester, 'gaussian'),
153                          'ConnectLayers failed to pass the KS test.')
154
155        def test_2D_gaussian_z(self):
156            '''Z-test performed on total connection count'''
157            self.assertTrue(self.z_test(ConnectLayers2DTester, 'gaussian'),
158                          'ConnectLayers failed to pass the Z-test')
159
160        def test_2D_linear_shifted_ks(self):
161            '''KS test performed on source-target node distances'''
162            self.assertTrue(self.ks_test(ConnectLayers2DTester, 'linear',
163                                  source_pos=source_pos / 4., self.L / 4.),
164                          'ConnectLayers failed to pass the KS test.')
165
166        def test_2D_linear_shifted_z(self):
167            '''Z-test performed on total connection count'''
168            self.assertTrue(self.z_test(ConnectLayers2DTester, 'linear',
169                                  source_pos=source_pos / 4., self.L / 4.),
170                          'ConnectLayers failed to pass the KS test.')
171
172        def test_2D_linear_multithread_ks(self):
173            '''KS test performed on source-target node distances'''
174            self.assertTrue(self.ks_test(ConnectLayers2DTester, 'linear',
175                                      threads=4),
176                          'ConnectLayers failed to pass the KS test.')
177
178        def test_2D_linear_multithread_z(self):
179            '''Z-test performed on total connection count'''
180            self.assertTrue(self.z_test(ConnectLayers2DTester, 'linear',
181                                      threads=4),
182                          'ConnectLayers failed to pass the Z-test')
183
184        def test_2D_linear_multithread_z(self):
185            '''Z-test performed on total connection count'''
186            self.assertTrue(self.z_test(ConnectLayers2DTester, 'linear',
187                                      threads=4),
188                          'ConnectLayers failed to pass the Z-test')
189
190        def test_3D_constant_ks(self):
191            '''KS test performed on source-target node distances'''
192
193

```

```

194     self.assertTrue(self.ks_test(ConnectLayers3DTester, 'constant',
195         kernel_params=0.5),
196         'ConnectLayers failed to pass the KS test.')
197
198     def test_3D_constant_z(self):
199         '''Z-test performed on total connection count'''
200
201         self.assertTrue(self.z_test(ConnectLayers3DTester, 'constant',
202             kernel_params=0.5),
203             'ConnectLayers failed to pass the Z-test')
204
205     def test_3D_linear_ks(self):
206         '''KS test performed on source-target node distances'''
207
208         self.assertTrue(self.ks_test(ConnectLayers3DTester, 'linear'),
209             'ConnectLayers failed to pass the KS test.')
210
211     def test_3D_linear_z(self):
212         '''Z-test performed on total connection count'''
213
214         self.assertTrue(self.z_test(ConnectLayers3DTester, 'linear'),
215             'ConnectLayers failed to pass the Z-test')
216
217     def test_3D_exponential_ks(self):
218         '''KS test performed on source-target node distances'''
219
220         self.assertTrue(self.ks_test(ConnectLayers3DTester, 'exponential'),
221             'ConnectLayers failed to pass the KS test.')
222
223     def test_3D_exponential_z(self):
224         '''Z-test performed on total connection count'''
225
226         self.assertTrue(self.z_test(ConnectLayers3DTester, 'exponential'),
227             'ConnectLayers failed to pass the Z-test')
228
229     def test_3D_gaussian_ks(self):
230         '''KS test performed on source-target node distances'''
231
232         self.assertTrue(self.ks_test(ConnectLayers3DTester, 'gaussian'),
233             'ConnectLayers failed to pass the KS test.')
234
235     def test_3D_gaussian_z(self):
236         '''Z-test performed on total connection count'''
237
238         self.assertTrue(self.z_test(ConnectLayers3DTester, 'gaussian'),
239             'ConnectLayers failed to pass the Z-test')
240
241     def test_3D_linear_shifted_ks(self):
242         '''KS test performed on source-target node distances'''
243
244         self.assertTrue(self.ks_test(ConnectLayers3DTester, 'linear',
245             source_pos=(self.L / 4., self.L / 4., self.L / 4.)),
246             'ConnectLayers failed to pass the KS test.')
247
248         def test_3D_linear_shifted_z(self):
249             '''Z-test performed on total connection count'''
250
251             self.assertTrue(self.z_test(ConnectLayers3DTester, 'linear',
252                 source_pos=(self.L / 4., self.L / 4., self.L / 4.)),
253                 'ConnectLayers failed to pass the Z-test')
254
255         def test_3D_linear_multithread_ks(self):
256             '''KS test performed on source-target node distances'''
257
258             self.assertTrue(self.ks_test(ConnectLayers3DTester, 'linear',
259                 threads=4),
260                 'ConnectLayers failed to pass the KS test.')
261
262         def test_3D_linear_multithread_z(self):
263             '''Z-test performed on total connection count'''
264
265             self.assertTrue(self.z_test(ConnectLayers3DTester, 'linear',
266                 threads=4),
267                 'ConnectLayers failed to pass the Z-test')
268
269         def suite():
270             suite = unittest.makeSuite(ConnectLayersTestCase, 'test')
271             return suite
272
273         if __name__ == '__main__':
274             runner = unittest.TextTestRunner(verbosity=2)
275             runner.run(suite())
276
277

```



# References

- Abbott, A. (2013). Brain-simulation and graphene projects win billion-euro competition. *Nature*.
- Brette, R. and W. Gerstner (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology* 94(5), 3637–3642.
- Brette, R., M. Rudolph, T. Carnevale, M. Hines, D. Beeman, J. M. Bower, M. Diesmann, A. Morrison, P. H. Goodman, F. C. H. Jr., M. Zirpe, T. Natschläger, D. Pecevski, B. Ermentrout, M. Djurfeldt, A. Lansner, O. Rochel, T. Vieville, E. Muller, A. P. Davison, S. E. Boustani, and A. Destexhe (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *J Comput Neurosci* 23, 349–398.
- Chang, G. and C. B. Roth (2001). Structure of MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* 293(5536), 1793–1800.
- Crook, S. M., J. A. Bednar, S. Berger, R. Cannon, A. P. Davison, M. Djurfeldt, J. Eppler, B. Kriener, S. Furber, B. Graham, et al. (2012). Creating, documenting and sharing network models. *Network: Computation in Neural Systems* 23(4), 131–149.
- Dijkstra, E. W., E. W. Dijkstra, and E. W. Dijkstra (1970). *Notes on structured programming*. The Netherlands: Technological University Eindhoven.
- Djurfeldt, M. (2012). The connection-set algebra—a novel formalism for the representation of connectivity structure in neuronal network models. *Neuroinformatics* 10(3), 287–304.
- Eppler, J. M., M. Helias, E. Muller, M. Diesmann, and M.-O. Gewaltig (2009). PyNEST: a convenient interface to the NEST simulator. *Frontiers in Neuroinformatics* 2(12).
- Ewens, W. and G. Grant (2004). *Statistical methods in bioinformatics: an introduction*, Volume 10. New York: Springer Science+Business Media.
- Gewaltig, M.-O. and M. Diesmann (2007). NEST (NEural Simulation Tool). *Scholarpedia* 2(4), 1430.

- Gewaltig, M.-O., A. Morrison, and H. E. Plesser (2012). NEST by example: An introduction to the neural simulation tool NEST. In N. Le Novère (Ed.), *Computational Systems Neurobiology*, pp. 533–558. Dordrecht: Springer Science+Business Media.
- Greenwood, P. and M. Nikulin (1996). *A guide to chi-squared testing*. New York: Wiley.
- Hill, S. and G. Tononi (2005). Modeling sleep and wakefulness in the thalamocortical system. *Journal of Neurophysiology* 93(3), 1671–1698.
- Huizinga, D. and A. Kolawa (2007). *Automated defect prevention: best practices in software management*. Hoboken, NJ: Wiley-IEEE Computer Society Press.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks* 14(6), 1569–1572.
- Kriener, B. (2012). Testing connectivity in spatial networks. Internal report.
- L’Ecuyer, P. (2004). Random number generation. In J. E. Gentle, W. Haerdle, and Y. Mori (Eds.), *Handbook of Computational Statistics*, pp. 35–70. Springer-Verlag.
- L’Ecuyer, P. and R. Simard (2007). TestU01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software* 33(4), 22.
- Miller, G. (2006). A scientist’s nightmare: Software problem leads to five retractions. *Science* 314(5807), 1856–1857.
- Nordlie, E. and H. E. Plesser (2010). Visualizing neuronal network connectivity with connectivity pattern tables. *Frontiers in Neuroinformatics* 3(39).
- Panik, M. J. (2005). *Advanced statistics from an elementary point of view*. Burlington, MA: Elsevier Academic Press.
- Plesser, H. E. and K. Austvoll (2009). Specification and generation of structured neuronal network models with the NEST Topology module. *BMC Neuroscience* 10, P56.
- Plesser, H. E. and H. Enger (2012). *NEST Topology User Manual for NEST 2.2*.
- Plesser, H. E., J. M. Eppler, A. Morrison, M. Diesmann, and M.-O. Gewaltig (2007). Efficient parallel simulation of large-scale neuronal networks on clusters of multiprocessor computers. In A.-M. Kermarrec, L. Bougé, and T. Priol (Eds.), *Euro-Par 2007 Parallel Processing*, Volume 4641 of *Lecture Notes in Computer Science*, pp. 672–681. Springer Berlin Heidelberg.

- The Human Brain Project Preparatory Study Consortium (2012). The Human Brain Project: A report to the European Commission. Lausanne. [http://www.humanbrainproject.eu/files/HBP\\_flagship.pdf](http://www.humanbrainproject.eu/files/HBP_flagship.pdf).
- Wang, Y. (1993). On the number of successes in independent trials. *Statistica Sinica* 3(2), 295–312.

