The Kauffman NK Model

A Stochastic Combinatorial Optimization Model for Complex Systems

Hemanshu Kaul

hkaul@math.uiuc.edu.

University of Illinois at Urbana-Champaign

Outline of the Talk

- Introduction
- Mathematical Description
- NK Model as a Stochastic Network
- Computational Strategies using Stochastic Networks
- Dependency Graph and Bounds on Order Statistics
- Analysis for underlying Normal Distribution
- Analysis for underlying Uniform Distribution
- Concentration of Measure

Introduction

We want to model systems composed of several interacting components, where each component can be in one of many possible states.

Objective : Maximize a measure of performance of the system based on contributions from each component, depending on the state of the component and its ‘interaction’ with its neighbors.

Background

In 1987, Kauffman and Levin introduced

The NK model

- N counts the number of components in the system
- K measures the ‘degree’ of interaction between components

Background

In 1987, Kauffman and Levin introduced

The NK model

- N counts the number of components in the system
- K measures the ‘degree’ of interaction between components

The NK model was originally proposed to study the evolution of genomes.

- system $\equiv$ genome
- components $\equiv$ genes
- states $\equiv$ gene mutations
- performance measure $\equiv$ fitness

Applications I

- In Biology
  - maturation of immune response
  - evolution of protein or RNA sequences
  - molecular quasi-species
- For example, an antibody (system) is a collection of amino acid sites (components) with each site containing one of twenty amino acids (states), then the affinity (performance measure) of an antibody for a particular antigen depends on how the chosen amino acids interact with each other.

Applications II

- In Physics and Management Science
  - spin glasses
  - effectiveness of a project team
  - process of organizational change

- For example, a spin glass is defined as a system consisting of contiguous atoms (components). For each atom, it is possible to select a spin up or spin down (states). The total energy (performance measure) depends on how the selected spins interact. The objective is to choose spins so that the energy is minimized.

Mathematical Description

System – A vector with N components, each of which can be in one of p possible states.
$\mathbf{x} = (x_0, ..., x_{N-1})$, with $x_i \in \{0, 1, 2, \ldots, p-1\}$ and the numbers $0, 1, 2, ..., p-1$ used as labels for the states.

Mathematical Description

System – A vector with N components, each of which can be in one of 2 possible states.
$\mathbf{x} = (x_0, ..., x_{N-1})$, with $x_i \in \{0, 1\}$ and the numbers $0, 1$ used as labels for the states.

Mathematical Description

System – A vector with N components, each of which can be in one of 2 possible states.
$\mathbf{x} = (x_0, ..., x_{N-1})$, with $x_i \in \{0, 1\}$ and the numbers $0, 1$ used as labels for the states.

Performance Measure –

$$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(\mathbf{x})$$

$\phi_i(\mathbf{x})$ is the performance contribution from each component i.

Performance Measure

Performance Measure –

$$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(\mathbf{x})$$

Performance Measure

Performance Measure –

$$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(\mathbf{x})$$

ϕ_i, the contribution of component i to the overall performance of the system depends on

- its own state, and
- the states of K 'neighboring' components.

Performance Measure

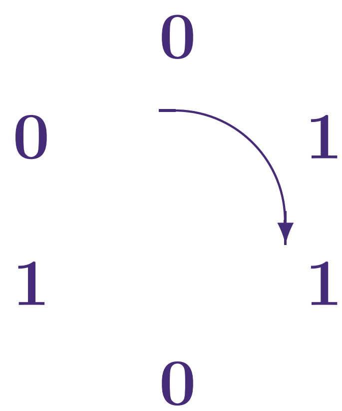
$N = 6$ and $K = 3$

System $(\mathbf{0}\ ,\ \mathbf{1}\ ,\ \mathbf{1}\ ,\ \mathbf{0}\ ,\ \mathbf{1}\ ,\ \mathbf{0})$

Performance Measure

$N = 6$ and $K = 3$

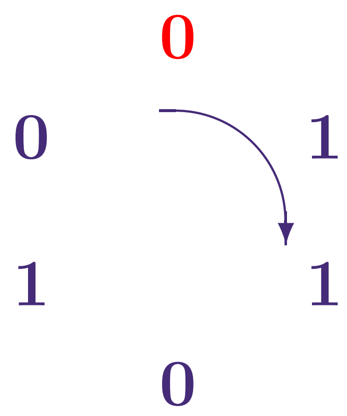
$$\mathbf{0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0}$$



Performance Measure

$N = 6$ and $K = 3$

0 1 1 0 1 0



Performance Measure

$N = 6$ and $K = 3$

$\mathbf{0} \quad \mathbf{1} \quad \mathbf{1} \quad \mathbf{0} \quad \mathbf{1} \quad \mathbf{0}$

$\mathbf{0}$

$\mathbf{0} \qquad \mathbf{1}$

$\mathbf{1} \qquad \mathbf{1}$

$\mathbf{0}$

Performance Measure

$N = 6$ and $K = 3$

$$0 \quad \color{red}{1} \quad 1 \quad 0 \quad 1 \quad 0$$

$$\begin{matrix} & 0 & \\ 0 & & \color{red}{1} \\ 1 & & 1 \\ & 0 & \end{matrix}$$

Performance Measure

$N = 6$ and $K = 3$

$$0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0$$

$$0$$

$$0 \qquad\qquad 1$$

$$1 \qquad\qquad 1$$

$$0$$

Performance Measure

$N = 6$ and $K = 3$

$$0 \quad 1 \quad 1 \quad 0 \quad \color{red}{1} \quad 0$$

$$0$$

$$0 \qquad\qquad 1$$

$$\color{red}{1} \qquad\qquad 1$$

$$0$$

Performance Measure

$N = 6$ and $K = 3$

$$0 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0$$

$$0$$

$$0 \qquad\qquad 1$$

$$1 \qquad\qquad 1$$

$$0$$

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0,1\}^{K+1}$.

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0,1\}^{K+1}$.

$N = 4$ and $K = 2$

$\Phi(\mathbf{0}, \mathbf{1}, \mathbf{1}, \mathbf{0}) =??$

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0,1\}^{K+1}$.

$N = 4$ and $K = 2$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1) + \ldots]$$

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0,1\}^{K+1}$.

$N = 4$ and $K = 2$

$$\Phi(\mathbf{0}, \mathbf{1}, \mathbf{1}, \mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,0) + \ldots]$$

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0, 1\}^{K+1}$.

$N = 4$ and $K = 2$

$$\Phi(\mathbf{0}, \mathbf{1}, \mathbf{1}, \mathbf{0}) = \frac{1}{4}[\phi_0(0, 1, 1) + \phi_1(1, 1, 0) + \phi_2(1, 0, 0) + \ldots]$$

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1}\phi_i(x_i,\ldots,x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0,1\}^{K+1}$.

$N = 4$ and $K = 2$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,0)+\phi_2(1,0,0)+\phi_3(0,0,1)]$$

Performance Measure - Definition, Example

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1}\phi_i(x_i,\ldots,x_{i+K})$$

where arithmetic in the subscripts is done modulo N and ϕ_i are N distinct real-valued functions on $\{0,1\}^{K+1}$.

$N = 4$ and $K = 2$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,0)+\phi_2(1,0,0)+\phi_3(0,0,1)]$$

Overlap

Question – Given N, K, $0 \leq K \leq N-1$, and $\phi_i : \{0,1\}^{K+1} \rightarrow \mathbb{R}$, $i = 0, 1, \ldots, N-1$

How can we find a system with the best possible performance ?

$$\max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$$

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,0)+\phi_1(0,0,0)+\phi_2(0,0,0)+\phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,1)+\phi_1(0,1,0)+\phi_2(1,0,0)+\phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,0)+\phi_2(1,0,0)+\phi_3(0,0,1)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,1)+\phi_2(1,1,0)+\phi_3(1,0,1)]$$

$$\Phi(\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(1,1,1)+\phi_1(1,1,1)+\phi_2(1,1,1)+\phi_3(1,1,1)]$$

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,0)+\phi_1(0,0,0)+\phi_2(0,0,0)+\phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,1)+\phi_1(0,1,0)+\phi_2(1,0,0)+\phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,0)+\phi_2(1,0,0)+\phi_3(0,0,1)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,1)+\phi_2(1,1,0)+\phi_3(1,0,1)]$$

$$\Phi(\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(1,1,1)+\phi_1(1,1,1)+\phi_2(1,1,1)+\phi_3(1,1,1)]$$

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,0) + \phi_1(0,0,0) + \phi_2(0,0,0) + \phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,1) + \phi_1(0,1,0) + \phi_2(1,0,0) + \phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,0) + \phi_2(1,0,0) + \phi_3(0,0,1)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,1) + \phi_2(1,1,0) + \phi_3(1,0,1)]$$

$$\Phi(\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(1,1,1) + \phi_1(1,1,1) + \phi_2(1,1,1) + \phi_3(1,1,1)]$$

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,0)+\phi_1(0,0,0)+\phi_2(0,0,0)+\phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,1)+\phi_1(0,1,0)+\phi_2(1,0,0)+\phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,0)+\phi_2(1,0,0)+\phi_3(0,0,1)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(0,1,1)+\phi_1(1,1,1)+\phi_2(1,1,0)+\phi_3(1,0,1)]$$

$$\Phi(\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(1,1,1)+\phi_1(1,1,1)+\phi_2(1,1,1)+\phi_3(1,1,1)]$$

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,0) + \phi_1(0,0,0) + \phi_2(0,0,0) + \phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,1) + \phi_1(0,1,0) + \phi_2(1,0,0) + \phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,0) + \phi_2(1,0,0) + \phi_3(0,0,1)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,1) + \phi_2(1,1,0) + \phi_3(1,0,1)]$$

$$\Phi(\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(1,1,1) + \phi_1(1,1,1) + \phi_2(1,1,1) + \phi_3(1,1,1)]$$

Central Question

Given N, K, $0 \leq K \leq N-1$, and
$\phi_i : \{0,1\}^{K+1} \to \mathbb{R}$, $i = 0, 1, \ldots, N-1$

What can we say about the Global Optima, the system that maximizes the value of the performance measure?

$$\max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$$

Central Question

Given N, K, $0 \leq K \leq N-1$, and
$\phi_i : \{0,1\}^{K+1} \to \mathbb{R}$, $i = 0, 1, \ldots, N-1$

What can we say about the Global Optima, the system that maximizes the value of the performance measure?

$$\max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$$

- NP-complete problem.
- In applications it is difficult, if not impossible, to determine the values taken by ϕ_i.

So, this combinatorial optimization problem is formulated and studied probabilistically.

Probability and Optimization I

Generate values of $\phi_i(.)$ stochastically.
For real-life scenarios in which the functions ϕ_i are not deterministically known, a universally adopted approach is to generate for each $\phi_i(.)$ a random number based on a probability distribution F.

This is analogous to replacing a “weight” in a combinatorial optimization model with a random variable, to better model uncertainty.

This is an idea inherent in Stochastic Programming.

Probability and Optimization II

“Average behavior” – Intractable combinatorial optimization problems are often studied probabilistically by introducing some notion of a random instance.

For example, in stochastic Traveling Salesman Problem (TSP), the distances (“weights”) between the vertices of a graph are replaced by $i.i.d$ uniform random variables. Replace $\phi_i(.)$ with random variables.

This is an idea inherent in Probabilistic Combinatorial Optimization.

Probabilistic Question

Given N, K, with $0 \le K \le N-1$, and $N2^{K+1}$ random variables $\phi_i(\mathbf{y})$ for $\mathbf{y} \in \{0,1\}^{K+1}$, $i = 0, 1, \ldots, N-1$, independently and identically distributed as F.

Study the distribution of the global optima –

$$\boldsymbol{X_{N,K} = \max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}}$$

where $\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$.

Overlap

$N = 4$ and $K = 2$

$2^N = 2^4 = 16$ possible systems

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,0) + \phi_1(0,0,0) + \phi_2(0,0,0) + \phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,0,1) + \phi_1(0,1,0) + \phi_2(1,0,0) + \phi_3(0,0,0)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,0) + \phi_2(1,0,0) + \phi_3(0,0,1)]$$

$$\Phi(\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(0,1,1) + \phi_1(1,1,1) + \phi_2(1,1,0) + \phi_3(1,0,1)]$$

$$\Phi(\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1}) = \frac{1}{4}[\phi_0(1,1,1) + \phi_1(1,1,1) + \phi_2(1,1,1) + \phi_3(1,1,1)]$$

Previous Research

Research Question– How do the varying values of N and K affect the performance of the systems?

Previous Research

Research Question– How do the varying values of N and K affect the performance of the systems?

- Mostly study of local optima w.r.t. a Hamming distance based neighborhood structure.
- Mostly simulation-based results and applications.
- Solow et. al (2000) showed the global decision problem is NP-complete.

Previous Research

- Evans and Steinsaltz (2002)
  - convert to an infinite-dimensional variational problem
  - explicit bounds only when $K=1$ and F is exponential distribution
- Durrett and Limic (2003)
  - use the theory of substochastic Harris chains
  - explicit bounds only when $K=1$ and F is negative exponential distribution
- Numerous other papers (both Applications and Theory).

Our Focus

Study $X_{N,K} = \max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$

- Develop a simple computational set-up, independent of the underlying distribution F

Our Focus

Study $X_{N,K} = \max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$

- Develop a simple computational set-up, independent of the underlying distribution F
- Develop methodology for finding bounds on the moments of $X_{N,K}$, independent of the underlying distribution F

Our Focus

Study $X_{N,K} = \max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$

- Develop a simple computational set-up, independent of the underlying distribution F
- Develop methodology for finding bounds on the moments of $X_{N,K}$, independent of the underlying distribution F
- Find explicit bounds on the expectation of $X_{N,K}$ when K is function of N, for fundamental underlying distributions like uniform and normal.

Our Focus

Study $X_{N,K} = \max\{\Phi(\mathbf{x}) \mid \mathbf{x} \in \{0,1\}^N\}$

- Develop a simple computational set-up, independent of the underlying distribution F
- Develop methodology for finding bounds on the moments of $X_{N,K}$, independent of the underlying distribution F
- Find explicit bounds on the expectation of $X_{N,K}$ when K is function of N, for fundamental underlying distributions like uniform and normal.
- Show concentration of $X_{N,K}$ around its mean, $\mathbf{E}_{N,K}$.

Our Focus

- Develop a simple computational set-up, independent of the underlying distribution F
- Develop methodology for finding bounds on the moments of $X_{N,K}$, independent of the underlying distribution F
- Find explicit bounds on the expectation of $X_{N,K}$ when K is function of N, for fundamental underlying distributions like uniform and normal.
- Show concentration of $X_{N,K}$ around its mean, $\mathbf{E}_{N,K}$.

We use tools from Combinatorics and Graph Theory, Networks, Probability and Statistics, and Geometry.

NK model as a Stochastic Network

Network $\boldsymbol{D_{N,K}}$ $2^{K+1} \times (N+1)$ array of vertices,

$v^i_{\mathbf{t}}$, $\mathbf{t} \in \{0,1\}^{K+1}$, $0 \leq i \leq N$

each vertex, $v^i_{\mathbf{t}}$, corresponds to component i and $\mathbf{t}$, the state vector for the component and its K neighbors.

NK model as a Stochastic Network

Network $\boldsymbol{D_{N,K}}$ $2^{K+1} \times (N+1)$ array of vertices,

$v^i_{\mathbf{t}}$, $\mathbf{t} \in \{0,1\}^{K+1}$, $0 \le i \le N$

each vertex, $v^i_{\mathbf{t}}$, corresponds to component i and $\mathbf{t}$, the state vector for the component and its K neighbors.

Idea – Create a correspondence between the systems and the directed paths in this network.

NK model as a Stochastic Network

Network $\boldsymbol{D_{N,K}}$ $\quad 2^{K+1} \times (N+1)$ array of vertices,

$v^i_{\mathbf{t}}$, $\mathbf{t} \in \{0,1\}^{K+1}$, $0 \leq i \leq N$

each vertex, $v^i_{\mathbf{t}}$, corresponds to component i and $\mathbf{t}$, the state vector for the component and its K neighbors.

Idea – Create a correspondence between the systems and the directed paths in this network.

$$v^i_{\mathbf{t}} \to v^j_{\hat{\mathbf{t}}} \quad \Leftrightarrow \quad j = i+1 \text{ and } \hat{t}_i = t_{i+1},\ i = 1, \ldots, K$$

$$\text{and } \hat{t}_{K+1} \in \{0,1\}$$

NK model as a Stochastic Network

Network $D_{N,K}$ $\quad 2^{K+1} \times (N+1)$ array of vertices,

$v^i_{\mathbf{t}}$, $\mathbf{t} \in \{0,1\}^{K+1}, \; 0 \leq i \leq N$

each vertex, $v^i_{\mathbf{t}}$, corresponds to component i and $\mathbf{t}$, the state vector for the component and its K neighbors.

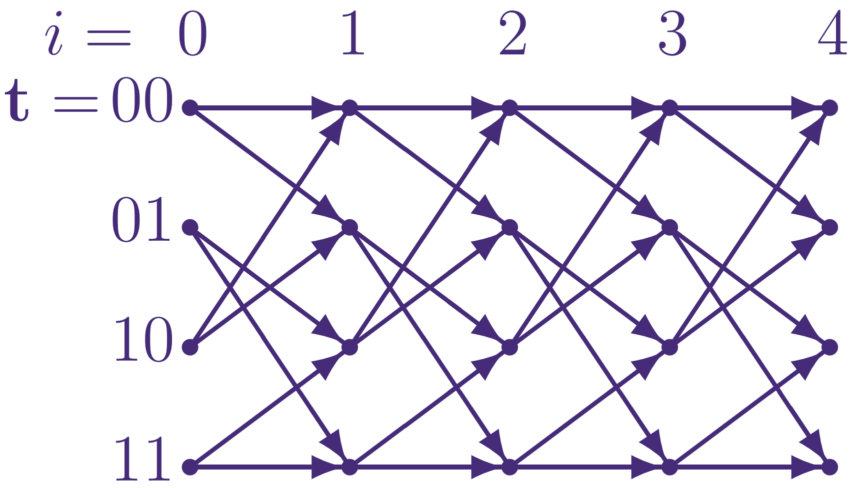
Idea – Create a correspondence between the systems and the directed paths in this network.

$$v^i_{\mathbf{t}} \rightarrow v^j_{\hat{\mathbf{t}}} \quad \Leftrightarrow \quad j = i + 1 \text{ and } \hat{t}_i = t_{i+1}, \; i = 1, \ldots, K$$
$$\text{and } \hat{t}_{K+1} \in \{0,1\}$$

Each $v^i_{\mathbf{t}}$ has a weight generated by the performance contribution (and random variable) $\phi_i(\mathbf{t})$.

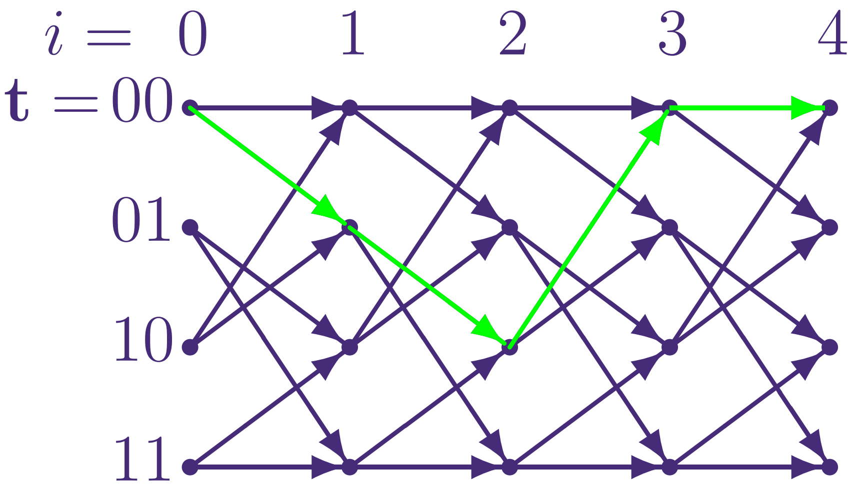
Network $D_{N,K}$

$\mathbf{N} = \mathbf{4}$ and $\mathbf{K} = \mathbf{1}$



Network $D_{N,K}$

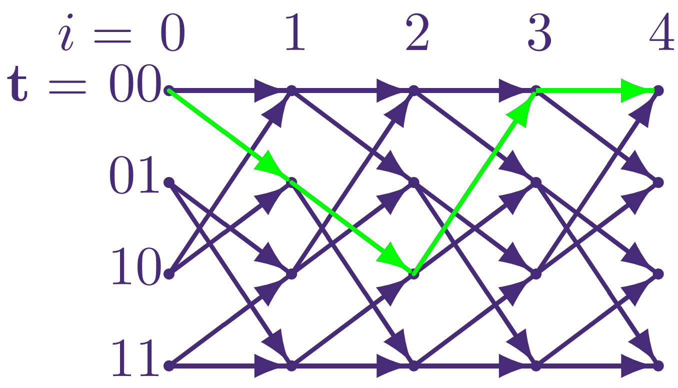
$\mathbf{N} = \mathbf{4}$ and $\mathbf{K} = \mathbf{1}$



Green path corresponds to the system $\{0, 0, 1, 0\}$ and the weight of the path is the performance measure of this system.

Network $D_{N,K}$

$\mathbf{N} = \mathbf{4}$ and $\mathbf{K} = \mathbf{1}$



Green path corresponds to the system $\{0, 0, 1, 0\}$ and the weight of the path is the performance measure of the system.

Each directed path from from $v_{\mathbf{t}}^{0}$ to $v_{\mathbf{t}}^{N}$

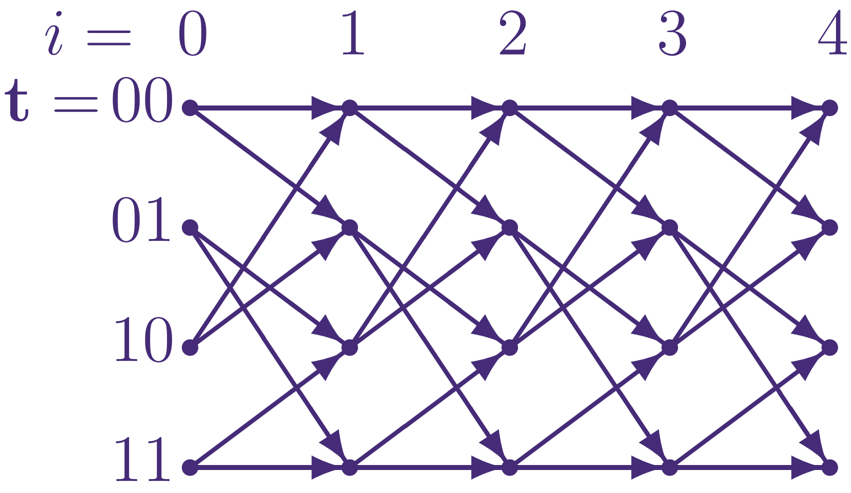
and its associated weight

$\updownarrow$

Each system and its performance

SubNetwork $D^{\mathbf{t}}_{N,K}$

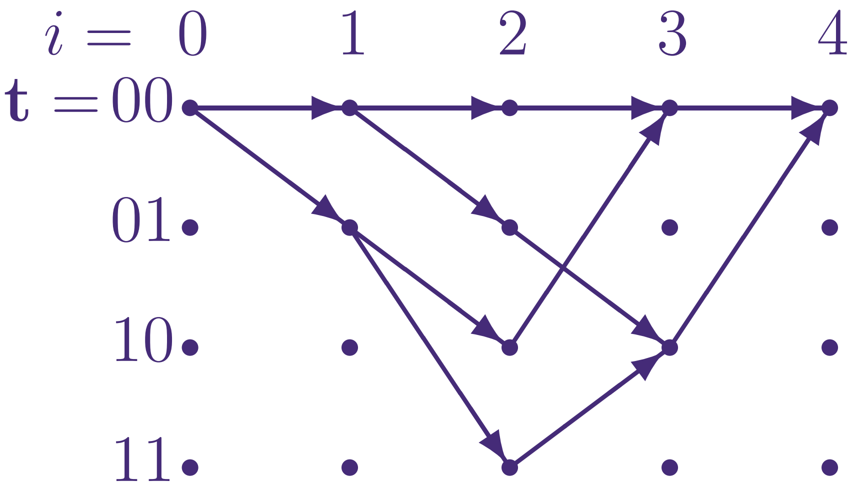
$D^{\mathbf{t}}_{N,K} \equiv$ subnetwork of $D_{N,K}$ defined by all the directed paths between $v^0_{\mathbf{t}}$ and $v^N_{\mathbf{t}}$



$D_{N,K}$

SubNetwork $D^{\mathbf{t}}_{N,K}$

$D^{\mathbf{t}}_{N,K} \equiv$ subnetwork of $D_{N,K}$ defined by all the directed paths between $v^0_{\mathbf{t}}$ and $v^N_{\mathbf{t}}$



$\boldsymbol{D^{0}_{N,K}}$

SubNetwork $D^{\mathbf{t}}_{N,K}$

$D^{\mathbf{t}}_{N,K} \equiv$ subnetwork of $D_{N,K}$ defined by all the directed paths between $v^0_{\mathbf{t}}$ and $v^N_{\mathbf{t}}$

$l^{\mathbf{t}}_{N,K} \equiv r.v.$ for maximum weight of a directed path in $D^{\mathbf{t}}_{N,K}$

SubNetwork $D^{\mathbf{t}}_{N,K}$

$D^{\mathbf{t}}_{N,K} \equiv$ subnetwork of $D_{N,K}$ defined by all the directed paths between $v^0_{\mathbf{t}}$ and $v^N_{\mathbf{t}}$

$l^{\mathbf{t}}_{N,K} \equiv r.v.$ for maximum weight of a directed path in $D^{\mathbf{t}}_{N,K}$

Since each of the 2^{K+1} subnetworks has identical structure, each $l^{\mathbf{t}}_{N,K}$ is identically distributed.

$l_{N,K} \equiv$ common $r.v.$ for each $l^{\mathbf{t}}_{N,K}$

SubNetwork $D^{\mathbf{t}}_{N,K}$

$D^{\mathbf{t}}_{N,K} \equiv$ subnetwork of $D_{N,K}$ defined by all the directed paths between $v^0_{\mathbf{t}}$ and $v^N_{\mathbf{t}}$

$l^{\mathbf{t}}_{N,K} \equiv r.v.$ for maximum weight of a directed path in $D^{\mathbf{t}}_{N,K}$

Since each of the 2^{K+1} subnetworks has identical structure, each $l^{\mathbf{t}}_{N,K}$ is identically distributed.

$l_{N,K} \equiv$ common $r.v.$ for each $l^{\mathbf{t}}_{N,K}$

$$\therefore X_{N,K} = \frac{1}{N}\max\{2^{K+1}\ \textit{identically distributed}\ \ l_{N,K}\}$$

SubNetwork $D^{\mathbf{t}}_{N,K}$

$D^{\mathbf{t}}_{N,K} \equiv$ subnetwork of $D_{N,K}$ defined by all the directed paths between $v^0_{\mathbf{t}}$ and $v^N_{\mathbf{t}}$

$l^{\mathbf{t}}_{N,K} \equiv r.v.$ for maximum weight of a directed path in $D^{\mathbf{t}}_{N,K}$

$l_{N,K} \equiv$ common $r.v.$ for each $l^{\mathbf{t}}_{N,K}$

$$X_{N,K} = \frac{1}{N}\max\{2^{K+1}\ \textit{identically distributed}\ \ l_{N,K}\ \}$$

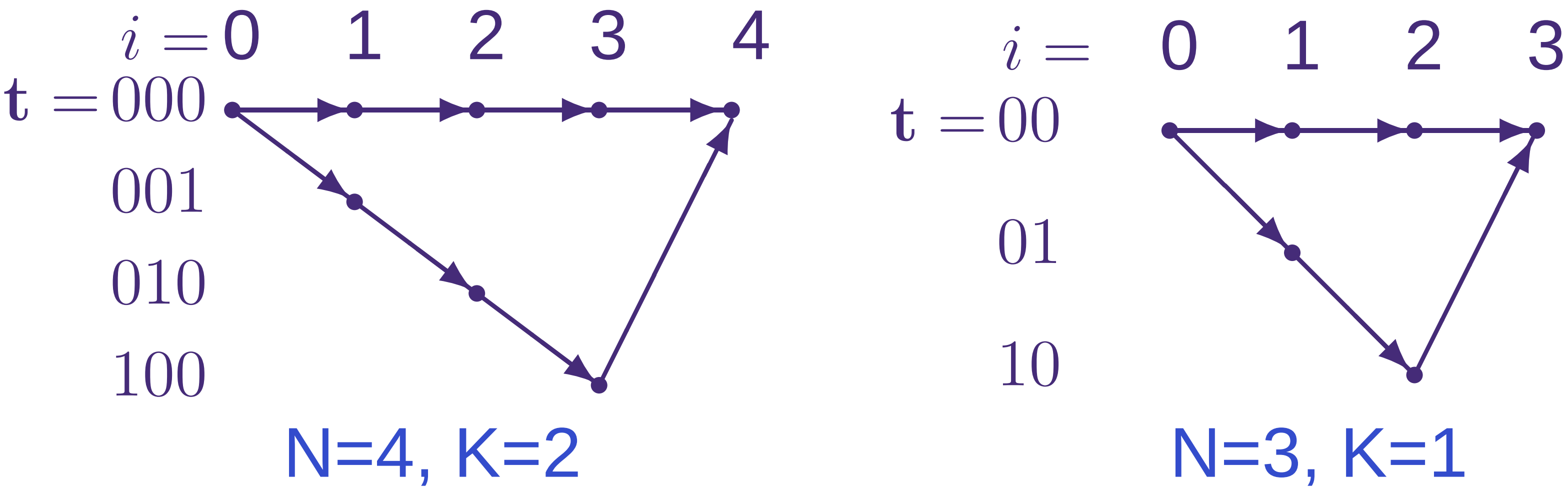
$$X_{N,K} = \frac{1}{N}\max\{2^{N}\ \textit{identically distributed}\ \ \Phi(\mathbf{x})\ \}$$

– Order Statistics

– Project Duration in PERT networks

Computational Strategy for K close to N

Observation – The value of $N-K$ determines the general structure of subnetwork $D^{\mathbf{t}}_{N,K}$, while N determines its size.



Subnetwork $D^{\mathbf{0}}_{N,K}$ for $N-K=2$

Computational Strategy for K close to N

This leads to –

For each K, $1 \le K \le N-3$,

$$l_{N,K} \quad = \quad X + \max\{two\ identically\ distributed\ l_{N-1,K}\}\ ,$$

where the boundary conditions are

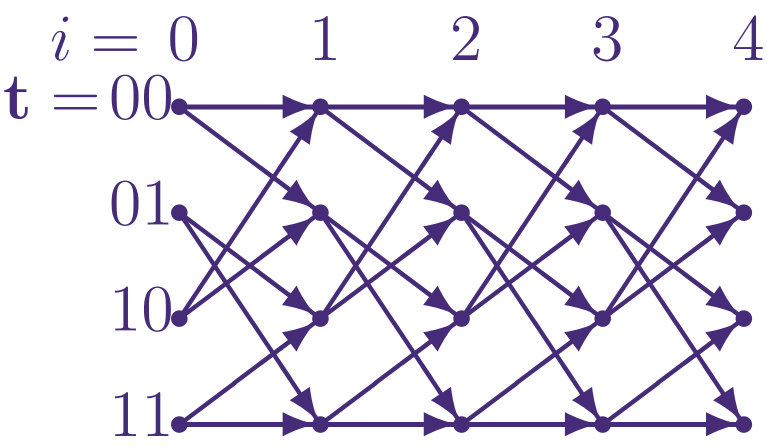
$$l_{K+2,K} \quad = \quad X + \max\{two\ \ i.i.d.\ \ l_{K+1,K}\}\ , \qquad X \sim F$$

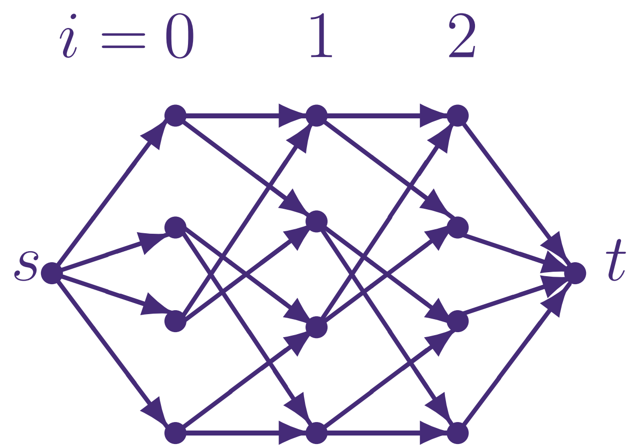
$$l_{K+1,K} \quad = \quad \sum_{i=1}^{N} X_i\ , \qquad \{X_i\}\ i.i.d.\ F$$

Each recursive step reduces the value of N and brings it closer to the (fixed) value of K, until $N = K + 1$.

$D'_{N,K}$ – Computational Strategy for small K

$D'_{N,K} \equiv$ Network formed from $D_{N,K}$ by deleting the vertices in the $K+1$ columns from $N-K$ to N and adding a source and a sink





$N=4\ ,\ K=1$ $D_{N,K}$ $D'_{N,K}$

$D'_{N,K}$ – Computational Strategy for small K

$D'_{N,K} \equiv$ Network formed from $D_{N,K}$ by deleting the vertices in the $K+1$ columns from $N-K$ to N and adding a source and a sink

Each directed path in $D'_{N,K}$ corresponds to a unique system, but not all feasible systems are represented by a path in $D'_{N,K}$.

$D'_{N,K}$ – Computational Strategy for small K

$D'_{N,K} \equiv$ Network formed from $D_{N,K}$ by deleting the vertices in the $K+1$ columns from $N-K$ to N and adding a source and a sink

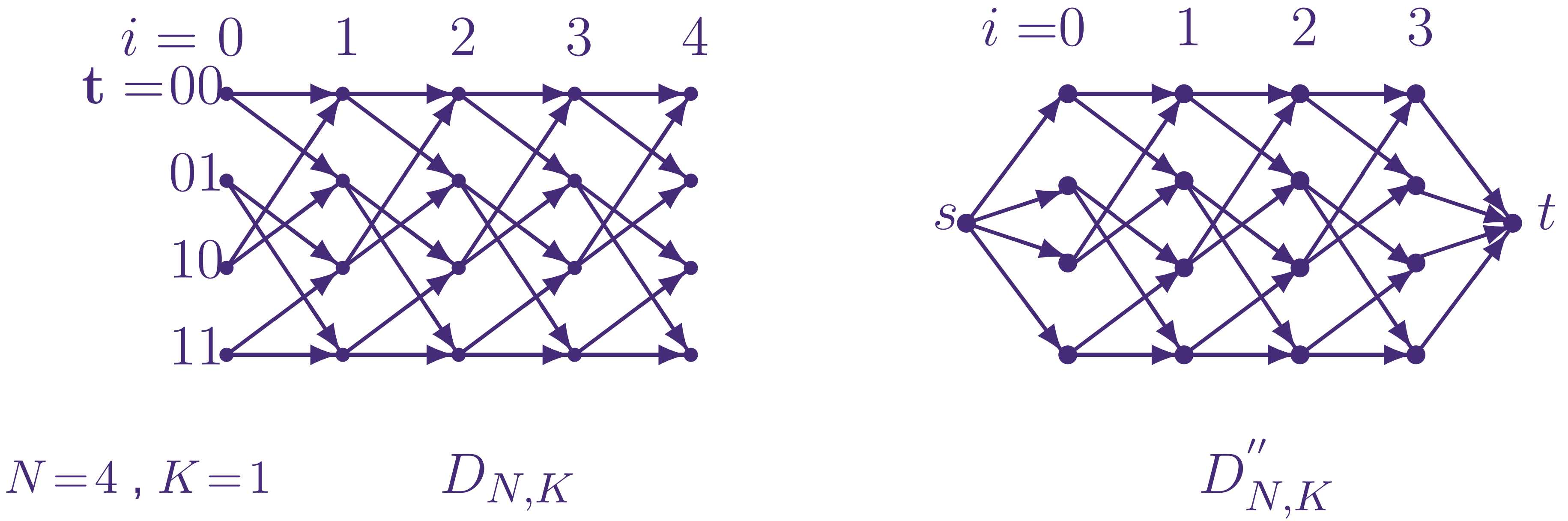
Each directed path in $D'_{N,K}$ corresponds to a unique system, but not all feasible systems are represented by a path in $D'_{N,K}$.

$$X_{N,K} \geq \frac{1}{N}\left[l'_{N,K} + \sum_{i=N-K}^{N-1} X_i\right], \quad X_i \; i.i.d. \; F$$

$l'_{N,K} \equiv$ maximum weight of a directed path in $D'_{N,K}$

$D''_{N,K}$ – Computational Strategy for small K

$D''_{N,K} \equiv$ Network formed from $D_{N,K}$ by deleting the vertices in column N and adding a source and a sink



$D''_{N,K}$ – Computational Strategy for small K

$D''_{N,K} \equiv$ Network formed from $D_{N,K}$ by deleting the vertices in column N and adding a source and a sink

Each feasible system corresponds to a unique directed path in $D''_{N,K}$, but not all directed paths represent a system.

$D''_{N,K}$ – Computational Strategy for small K

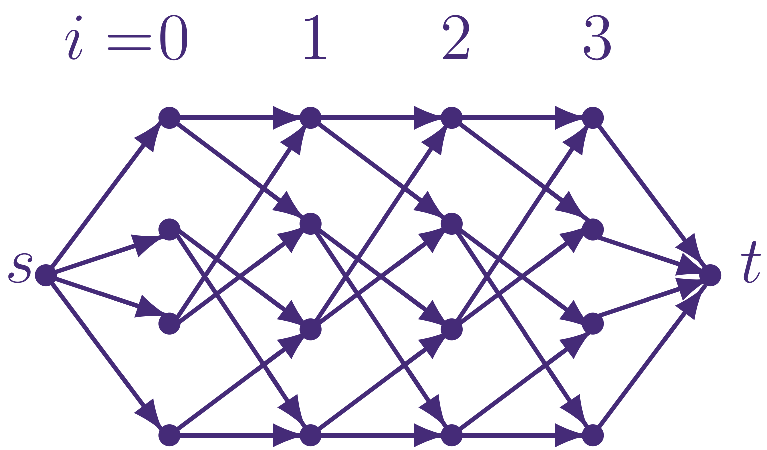
$D''_{N,K} \equiv$ Network formed from $D_{N,K}$ by deleting the vertices in column N and adding a source and a sink

Each feasible system corresponds to a unique directed path in $D''_{N,K}$, but not all directed paths represent a system.

$$X_{N,K} \leq \frac{1}{N} \left[l''_{N,K} \right]$$

$l''_{N,K} \equiv$ maximum weight of a directed path in $D''_{N,K}$

$D'_{N,K}$ and $D''_{N,K}$



$N = 4$, $K = 1$ $D''_{N,K}$ $D'_{N,K}$

$D'_{N,K}$ has $N - K$ columns and $D''_{N,K}$ has N columns.

For fixed K, the bounds in terms of $l'_{N,K}$ and $l''_{N,K}$ will be asymptotically tight.

Dependency Graph

$$X_{N,K} = \frac{1}{N}\max\{2^N \textit{ identically distributed } \Phi(\mathbf{x})\}$$

Dependency Graph

$$X_{N,K} = \frac{1}{N}\max\{2^N \ \textit{identically distributed}\ \ \Phi(\mathbf{x})\ \}$$

Dependence between $\Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \{0,1\}^N$

$$\Phi(\mathbf{x}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$$

$$\Phi(\mathbf{y}) = \frac{1}{N}\sum_{i=0}^{N-1} \phi_i(y_i, \ldots, y_{i+K})$$

Dependency Graph

$X_{N,K} = \frac{1}{N} \max\{2^N \; identically \; distributed \;\; \Phi(\mathbf{x}) \}$

Dependence between $\Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \{0,1\}^N$

$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$

$\Phi(\mathbf{y}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(y_i, \ldots, y_{i+K})$

$\Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$ are dependent $\Leftrightarrow$
there exists i such that $x_j = y_j$ for $i \leq j \leq i + K$

Dependency Graph

Dependence between $\Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$, $\mathbf{x}, \mathbf{y} \in \{0,1\}^N$

$\Phi(\mathbf{x}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(x_i, \ldots, x_{i+K})$

$\Phi(\mathbf{y}) = \frac{1}{N} \sum_{i=0}^{N-1} \phi_i(y_i, \ldots, y_{i+K})$

$\Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$ are dependent $\Leftrightarrow$
there exists i such that $x_j = y_j$ for $i \leq j \leq i+K$

$G_{N,K} \equiv$ dependency graph for given N, K

vertices $\equiv$ $\mathbf{x} \in \{0,1\}^N$

$\mathbf{x} \leftrightarrow \mathbf{y} \Leftrightarrow \Phi(\mathbf{x})$ and $\Phi(\mathbf{y})$ are dependent

Dependency Graph, contd.

- $G_{N,K}$ has 2^N vertices, one for each system.
- An edge between two vertices means there is dependence between the performance measures of the corresponding systems.

Dependency Graph, contd.

- $G_{N,K}$ has 2^N vertices, one for each system.
- An edge between two vertices means there is dependence between the performance measures of the corresponding systems.

- Want to partition the vertex set of $G_{N,K}$,
  $V(G_{N,K}) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_t$, such that

  - there are no edges within each class V_i

  - sizes of any two classes differ by at most 1

Dependency Graph, contd.

Want to partition the vertex set of $G_{N,K}$, $V(G_{N,K}) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_t$, such that

- there are no edges within each class V_i

- sizes of any two classes differ by at most 1

t-equitable coloring of $G_{N,K}$

Dependency Graph, contd.

Want to partition the vertex set of $G_{N,K}$,
$V(G_{N,K}) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_t$, such that

- there are no edges within each class V_i

- sizes of any two classes differ by at most 1

t-equitable coloring of $G_{N,K}$

Theorem : $\Delta(G_{N,K}) \leq N2^{N-K-2}$ for all K, with equality for $\frac{N}{2} \leq K \leq N-2$.

$\Delta(G) \equiv$ maximum degree, the most number of vertices that are adjacent to a vertex in G

Dependency Graph, contd.

Want to partition the vertex set of $G_{N,K}$,
$V(G_{N,K}) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_t$, such that

- there are no edges within each class V_i

- sizes of any two classes differ by at most 1

t-equitable coloring of $G_{N,K}$

Theorem : $\Delta(G_{N,K}) \leq N2^{N-K-2}$ for all K, with equality for $\frac{N}{2} \leq K \leq N-2$.

Theorem : $G_{N,K}$ has a t-equitable coloring if $t > N2^{N-K-2}$

Dependency Graph, contd.

Want to partition the vertex set of $G_{N,K}$,
$V(G_{N,K}) = V_1 \sqcup V_2 \sqcup \ldots \sqcup V_t$, such that

- there are no edges within each class V_i

- sizes of any two classes differ by at most 1

t-equitable coloring of $G_{N,K}$

Theorem : $\Delta(G_{N,K}) \leq N2^{N-K-2}$ for all K, with equality for $\frac{N}{2} \leq K \leq N-2$.

Theorem : $G_{N,K}$ has a t-equitable coloring if $t > N2^{N-K-2}$

How is this useful?

Bounds on Order Statistics

Notation : $Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$

$F_N \equiv$ distribution of $\sum_{i=1}^{N} X_i$, for $X_i \; i.i.d. \; F$

Bounds on Order Statistics

Notation : $Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$

$F_N \equiv$ distribution of $\sum_{i=1}^{N} X_i$, for X_i *i.i.d.* F

$$\begin{aligned}
\mathbf{X_{N,K}} &= \frac{1}{N}\max\{2^N \textit{ identically distributed } \Phi(\mathbf{x})\,\} \\
&= \frac{1}{N}\max\{2^N \textit{ identically distributed } \sum_{i=1}^{N}\phi_i\,\}\,,\ \{\phi_i\}\ i.\ i.\ d.\ F \\
&= \frac{1}{N}\max\{2^N \textit{ identically distributed } \Phi(\mathbf{x})\,\}\,,\ \Phi(\mathbf{x}) \sim F_N \\
&= \frac{1}{N}\,Y_{[2^N]}\,,\ Y_i \sim F_N\ ;\{Y_i\,|\,i=1,\ldots,2^N\} = \{\Phi(\mathbf{x})\,|\,x \in \{0,1\}^N\}
\end{aligned}$$

Bounds on Order Statistics

Notation : $Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$

$F_N \equiv$ distribution of $\sum_{i=1}^{N} X_i$, for X_i $i.i.d.$ F

$\boldsymbol{X_{N,K}} = \frac{\mathbf{1}}{\boldsymbol{N}} \boldsymbol{Y_{[2^N]}}$, $Y_i \sim F_N$; $\{Y_i\} = \{\Phi(\mathbf{x})\}$ dependent

Bounds on Order Statistics

Notation : $Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$

$F_N \equiv$ distribution of $\sum_{i=1}^{N} X_i$, for $X_i \; i.i.d. \; F$

$\boldsymbol{X_{N,K}} = \frac{\mathbf{1}}{\boldsymbol{N}} \boldsymbol{Y}_{[\mathbf{2}^{\boldsymbol{N}}]}$, $Y_i \sim F_N$; $\{Y_i\} = \{\Phi(\mathbf{x})\}$ dependent

Theorem : For all N, K, with underlying distribution F, if $G_{N,K}$ has t-equitable coloring then

$$\mathbf{E}[Y_{[2^N/t]}] \leq \mathbf{E}[X_{N,K}] \leq \mathbf{E}[Y_{[2^N/t]}] + \sqrt{t \ \mathbf{Var}[Y_{[2^N/t]}]}$$

where $Y_1, \ldots, Y_k$ i.i.d. F_N.

Bounds on Order Statistics

Notation : $Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$

$F_N \equiv$ distribution of $\sum_{i=1}^{N} X_i$, for $X_i \; i.i.d. \; F$

$\boldsymbol{X_{N,K}} = \frac{\mathbf{1}}{\boldsymbol{N}} \, \mathbf{Y}_{[\mathbf{2}^{N}]}$, $Y_i \sim F_N$; $\{Y_i\} = \{\Phi(\mathbf{x})\}$ dependent

Theorem : For all N, K, with underlying distribution F,

$$\mathbf{E}[Y_{[2^{K+2}/N]}] \leq \mathbf{E}[X_{N,K}] \leq \mathbf{E}[Y_{[2^{K+2}/N]}] + \sqrt{N 2^{N-K-2} \mathbf{Var}[Y_{[2^{K+2}/N]}]}$$

where $Y_1, \ldots, Y_k$ **i.i.d.** F_N.

Bounds on Order Statistics

Notation : $Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$

$F_N \equiv$ distribution of $\sum_{i=1}^{N} X_i$, for $X_i\ i.i.d.\ F$

$\boldsymbol{X_{N,K} = \frac{1}{N}\, Y_{[2^N]}}$, $Y_i \sim F_N$; $\{Y_i\} = \{\Phi(\mathbf{x})\}$ dependent

Theorem : For all N, K, with underlying distribution F,

$$\mathbf{E}[Y_{[2^{K+2}/N]}] \leq \mathbf{E}[X_{N,K}] \leq \mathbf{E}[Y_{[2^{K+2}/N]}] + \sqrt{N 2^{N-K-2} \mathbf{Var}[Y_{[2^{K+2}/N]}]}$$

where $Y_1, \ldots, Y_k$ **i.i.d.** F_N.

$Proofs$ use tools from Order Statistics and the Equitable Coloring of Graphs.

Order Statistics with Dependencies

A *dependency graph* for random variables $X_1, \ldots, X_n$, $G(X_1, \ldots, X_n)$, has vertex set $[n]$ and an edge set such that for each $i \in [n]$, X_i is mutually independent of all other X_j such that $\{i, j\}$ is not an edge.

$$Y_{[n]} = \max\{Y_1, \ldots, Y_n\}$$

Theorem : Let $X_1, \ldots, X_n$ be identically distributed random variables with distribution F. If $G(X_1, \ldots, X_n)$ has a t-equitable coloring, then

$$\mathbf{E}[Y_{[n/t]}] \leq \mathbf{E}[X_{[n]}] \leq \mathbf{E}[Y_{[n/t]}] + \sqrt{(t-1)\,\mathbf{Var}[Y_{[n/t]}]}$$

where $Y_1, \ldots, Y_k$ $i.i.d.$ F .

Order Statistics with Dependencies

Theorem : Let $X_1, \ldots, X_n$ be identically distributed (dependent) random variables with distribution F. If $G(X_1, \ldots, X_n)$ has a t-equitable coloring, then

$$\mathbf{E}[Y_{[n/t]}] \leq \mathbf{E}[X_{[n]}] \leq \mathbf{E}[Y_{[n/t]}] + \sqrt{(t-1)\,\mathbf{Var}[Y_{[n/t]}]}$$

where $Y_1, \ldots, Y_k$ ***i.i.d.*** F .

Convert the problem of bounding order statistics of dependent random variables into that of independent random variables while incorporating quantitative information about the mutual dependencies between the original random variables

Bounds when $F = \mathbf{n}(0,1)$

Theorem :For all $N \geq 2$, $K = N - 1$,

$$\sqrt{2\log 2} - \frac{o(1)}{\sqrt{N}} \;\leq\; \mathbf{E}[X_{N,K}] \;\leq\; \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Bounds when $F = \mathbf{n}(0,1)$

Theorem :For all $N \geq 2$, $K = N - 1$,

$$\sqrt{2\log 2} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem :For all $N \geq 2$, $K = N - \alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\sqrt{(1-\frac{c}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Bounds when $F = \mathbf{n}(0,1)$

Theorem :For all $N \geq 2$, $K = N - 1$,

$$\sqrt{2\log 2} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1 + \frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem :For all $N \geq 2$, $K = N - \alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\sqrt{(1 - \frac{c}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1 + \frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\sqrt{(\beta + \frac{2}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1 + \frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Bounds when $F = \mathbf{n}(0,1)$

Theorem :For all $N \geq 2$, $K = N - 1$,

$$\sqrt{2\log 2} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem :For all $N \geq 2$, $K = N - \alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\sqrt{(1-\frac{c}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\sqrt{(\beta+\frac{2}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Tight bounds on $\mathbf{E}[X_{N,K}]$ valid for all N and for K close to N

Bounds when $F = \mathbf{n}(0,1)$

Theorem : For all $N \geq 2$, $K = N - 1$,

$$\sqrt{2\log 2} - \frac{o(1)}{\sqrt{N}} \;\leq\; \mathbf{E}[X_{N,K}] \;\leq\; \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem : For all $N \geq 2$, $K = N - \alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\sqrt{(1-\frac{c}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\sqrt{(\beta+\frac{2}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Leading Coefficients in both upper and lower bounds are equal to $\sqrt{2\log 2}$

Bounds when $F = \mathbf{n}(0,1)$

Theorem :For all $N \geq 2$, $K = N - 1$,

$$\sqrt{2\log 2} - \frac{o(1)}{\sqrt{N}} \;\leq\; \mathbf{E}[X_{N,K}] \;\leq\; \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem :For all $N \geq 2$, $K = N-\alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\sqrt{(1-\frac{c}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\sqrt{(\beta+\frac{2}{N})2\log 2 - \frac{2\log N}{N}} - \frac{o(1)}{\sqrt{N}} \leq \mathbf{E}[X_{N,K}] \leq \sqrt{(1+\frac{1}{N})2\log 2} - \frac{o(1)}{\sqrt{N}}$$

$Proofs$ use the previous Theorems and the properties of Normal Distribution & its order statistics

Bounds when $F = \mathbf{u}(0, 1)$

- "Sum of Normals is Normal"!
- Sum of Uniforms does not have a nice distribution.

Need to find an alternate description of the Distribution of sum of Uniforms !

Bounds when $F = \mathbf{U}(0,1)$

When $\{X_j\}$ $i.i.d.$ $\mathbf{U}(0,1)$,

$\mathbf{Pr}\left\{\sum_{j=1}^{N} X_j \le x\right\}$ is equal to the volume of

$$P(x) = \left\{\mathbf{y} \in \mathbb{R}^N \mid \sum_{j=1}^{N} y_j \le x \text{ and } 0 \le y_j \le 1\right\}$$

a subset of the N-dimensional hypercube $[0,1]^N$.

Bounds when $F = \mathbf{u}(0, 1)$

We prove lemmas about $Vol(P(x))$ that help to decompose the expectation integral.

Bounds when $F = \mathbf{u}(0,1)$

We prove lemmas about $Vol(P(x))$ that help to decompose the expectation integral.

For Example,
a lower bound on x that forces the volume of $P(x)$ to approach 1, the volume of the $[0,1]^N$ cube, very rapidly.

Lemma :

If $\boldsymbol{x > (1 - \frac{1}{2e})N}$, then $\boldsymbol{Vol(P(x)) \geq 1 - \frac{1}{\sqrt{2\pi N}\, 2^N}}$
for all $N \geq 2$.

Bounds when $F = \mathbf{u}(0,1)$

We prove lemmas about $Vol(P(x))$ that help to decompose the expectation integral.

For Example,
if the volume outside $P(x)$ is asymptotically small then x must be sufficiently large.

Lemma :
If $\boldsymbol{Vol(P(x)) > 1 - \frac{N}{2^N}}$, then $\boldsymbol{x > \left(1 - \frac{1}{4}(2N)^{1/N}\right)N}$ for all $N \geq 2$.

Bounds when $F = \mathbf{u}(0,1)$

Theorem : For all $N \geq 2$, $K = N-1$,

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{2^N}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\frac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$$\lim_{N\to\infty} \mathbf{Var}[X_{N,K}] \leq \frac{7}{16}-\frac{1}{e}\left(1-\frac{1}{2e}\right) \approx 0.1373$$

Bounds when $F = \mathbf{u}(0,1)$

Theorem : For all $N \geq 2$, $K = N - 1$,

$$\left(1 - \tfrac{1}{4}(2N)^{1/N}\right)\left(1 - \left(1 - \tfrac{N}{2^N}\right)^{2^N}\right) \leq \mathbf{E}[X_{N,K}] \leq 1 - \tfrac{1}{2e}\left(1 - \frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$$\lim_{N\to\infty} \mathbf{Var}[X_{N,K}] \leq \tfrac{7}{16} - \tfrac{1}{e}\left(1 - \tfrac{1}{2e}\right) \approx 0.1373$$

Theorem : For all $N \geq 2$, $K = N - \alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\left(1 - \tfrac{1}{4}(2N)^{1/N}\right)\left(1 - \left(1 - \tfrac{N}{2^N}\right)^{\frac{2^N}{cN}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1 - \tfrac{1}{2e}\left(1 - \frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Bounds when $F = \mathbf{u}(0,1)$

Theorem : For all $N \geq 2$, $K = N - 1$,

$$\left(1 - \tfrac{1}{4}(2N)^{1/N}\right)\left(1 - \left(1 - \tfrac{N}{2^N}\right)^{2^N}\right) \leq \mathbf{E}[X_{N,K}] \leq 1 - \tfrac{1}{2e}\left(1 - \frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$$\lim_{N\to\infty} \mathbf{Var}[X_{N,K}] \leq \tfrac{7}{16} - \tfrac{1}{e}\left(1 - \tfrac{1}{2e}\right) \approx 0.1373$$

Theorem : For all $N \geq 2$, $K = N - \alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\left(1 - \tfrac{1}{4}(2N)^{1/N}\right)\left(1 - \left(1 - \tfrac{N}{2^N}\right)^{\frac{2^N}{cN}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1 - \tfrac{1}{2e}\left(1 - \frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\left(1 - \tfrac{1}{4}(2N)^{1/N}\right)\left(1 - \left(1 - \tfrac{N}{2^N}\right)^{4\frac{2^{\beta N}}{N}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1 - \tfrac{1}{2e}\left(1 - \frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Bounds when $F = \mathbf{u}(0,1)$

Theorem : For all $N \geq 2$, $K = N-1$,

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{2^N}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$$\lim_{N\to\infty} \mathbf{Var}[X_{N,K}] \leq \tfrac{7}{16}-\tfrac{1}{e}\left(1-\tfrac{1}{2e}\right) \approx 0.1373$$

Theorem : For all $N \geq 2$, $K = N-\alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{\frac{2^N}{cN}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{4\frac{2^{\beta N}}{N}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Tight bounds on $\mathbf{E}[X_{N,K}]$ valid for all N and for K close to N

Bounds when $F = \mathbf{U}(0,1)$

Theorem : For all $N \geq 2$, $K = N - 1$,

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{2^N}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$$\lim_{N\to\infty} \mathbf{Var}[X_{N,K}] \leq \tfrac{7}{16}-\tfrac{1}{e}\left(1-\tfrac{1}{2e}\right) \approx 0.1373$$

Theorem : For all $N \geq 2$, $K = N-\alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{\frac{2^N}{cN}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{4\frac{2^{\beta N}}{N}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Leading Coefficients : $1-\frac{1}{4} = 0.75$ and $1-\frac{1}{2e} \approx 0.816$

Bounds when $F = \mathbf{U}(0,1)$

Theorem : For all $N \geq 2$, $K = N-1$,

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{2^N}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$$\lim_{N\to\infty} \mathbf{Var}[X_{N,K}] \leq \tfrac{7}{16}-\tfrac{1}{e}\left(1-\tfrac{1}{2e}\right) \approx 0.1373$$

Theorem : For all $N \geq 2$, $K = N-\alpha$, $\alpha \in \mathbb{Z}^+$, $\alpha \geq 2$, $c = \alpha - 2$

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{\frac{2^N}{cN}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

Theorem : For all $N \geq 2$, $K = \beta N$, $\beta \in (0,1)$

$$\left(1-\tfrac{1}{4}(2N)^{1/N}\right)\left(1-\left(1-\tfrac{N}{2^N}\right)^{4\frac{2^{\beta N}}{N}}\right) \leq \mathbf{E}[X_{N,K}] \leq 1-\tfrac{1}{2e}\left(1-\frac{1}{\sqrt{2\pi N}\,2^N}\right)^{2^N}$$

$Proofs$ use the previous Theorems and the geometric lemmas .

Concentration of $X_{N,K}$ around $\mathbf{E}[X_{N,K}]$

Probability of $X_{N,K}$ being far from $\mathbf{E}[X_{N,K}]$ is exponentially decaying.

Theorem : If F is a bounded distribution such that $X \sim F \Rightarrow |X| \leq c$, then

$$\mathbf{P}[\,|\,X_{N,K} - \mathbf{E}[X_{N,K}]\,|\geq t\,] \leq 2\exp\left(-\frac{2Nt^2}{c^2 2^{2N-K-1}}\right)$$

Concentration of $X_{N,K}$ around $\mathbf{E}[X_{N,K}]$

Probability of $X_{N,K}$ being far from $\mathbf{E}[X_{N,K}]$ is exponentially decaying.

Theorem : If F is a bounded distribution such that $X \sim F \Rightarrow |X| \leq c$, then

$$\mathbf{P}[\,|\,X_{N,K} - \mathbf{E}[X_{N,K}]\,|\geq t\,] \leq 2\exp\left(-\frac{2Nt^2}{c^2 2^{2N-K-1}}\right)$$

Proof using Independent Bounded Differences Inequality, a variant of Azuma's Martingale inequality.

The Kauffman NK Model

- Background and Applications
- Mathematical Description
- NK Model as a Stochastic Network
- Computational Strategies using Stochastic Networks
- Dependency Graph and Bounds on Order Statistics
- Analysis for underlying Normal Distribution
- Analysis for underlying Uniform Distribution
- Concentration of Measure

Thank You !