

## ネットワーク上の相関のある二項分布

守 真 太 郎<sup>†1</sup> 久 門 正 人<sup>†2</sup>

本論文では、ネットワーク上の相関のある二項分布モデルを提案し、確率分布関数の振舞いについて報告する。条件付確率を用いて、隣接する確率変数を逐次連結し、一次元格子、ケーリー木、一般化ランダムグラフなどのネットワーク上の相関のある二項分布を構成し、確率分布関数のネットワーク構造依存性を調べる。また、スケールフリーネットワーク上の相関のある二項分布と Nath ら (2006) により導入された二重指数分布関数を比較し、インターネット上のサーバのダウン相関が非常に強いと考えられることも指摘する。

### Correlated Binomial Model on Networks

SHINTARO MORI<sup>†1</sup> and MASATO HISAKADO<sup>†2</sup>

We introduce correlated binomial models on networks and study their probability functions. Using conditional probabilities, we connect random variables iteratively and construct the joint probability functions of the models on networks, including one dimensional lattice, cayley tree, generic random graph and so on. We study the probability functions and their dependences on network structures. Comparing correlated binomial models on scale-free networks and the bi-exponential model introduced by Nath, et al. (2006), we show that the failure correlation between internet servers are very strong.

#### 1. はじめに

この世の中の事象は独立に起こることは稀であり、一般に互いに依存しながら生起する。

そのため、個々の事象が起こるか起こらないかという確率が同じであっても、どの程度の大きさの事象の連鎖が起こったのかという問題を考える際には、事象間の依存関係をモデルに組み込む必要がある。計量生物学の分野では、ねずみの胎児に対する毒物の効果や虫歯の本数などの分布を扱う際に、胎児の生死、虫歯になるかならないかが独立には起きないことから、ベータ二項分布が用いられてきた<sup>2)–5)</sup>。

金融工学の分野では、近年クレジットデリバティブの格付け、プライシングに関連して、企業や債務者のデフォルト（債務不履行）をモデル化する必要性が生じた。デフォルトは独立には起こらず、1つのデフォルトは他の債務者に直接的にも間接的にも影響を及ぼし、連鎖的に多数の債務者がデフォルトするリスクがある。クレジットリスクの分野では、ベータ二項分布ではなく、相関のある正規分布を用いて、デフォルト間の相関をとりいれるモデルが、ファクターモデル、正規コピュラモデルとして標準的なものとなっている<sup>6)</sup>。しかし、そこから計算されるクレジットデリバティブのプレミアムは、マーケットでの価格と乖離する。そのため、デリバティブごとに理論価格とマーケット価格をあわせたインプライド相関を使うか、または正規分布を放棄し、プレミアムからデフォルト分布を推定（インプライド分布と呼ばれる）して、それをもとに他のデリバティブのプレミアムを計算するといった、その場しのぎの手法が現在用いられている<sup>7)</sup>。サブプライム問題が発生した2007年の夏以降、CDOのマーケットは開店休業の事態に追い込まれたが、その原因の1つがデフォルト連鎖を記述する確固とした確率モデルの欠如<sup>\*1</sup>である。デフォルト連鎖を記述することが難しい理由は、どの債務者とどの債務者がどのような依存関係にあり、ある債務者がデフォルトしたとき、依存関係にある債務者のデフォルトにどの程度影響したのかの定量的なデータがないためである。

一方、情報工学の分野でも類似の問題を取り扱うことが必要になってきている。サーバの連鎖ダウンの問題がそれで、近年さまざまなネットワークのサーバのダウンについて、どの程度のものがどの程度の頻度で起こるのかについてデータを蓄積し、それをもとに安全なネットワークの設計指針が考えられてきた。特に Nath ら<sup>1)</sup> は3つの異なるサーバネットワークのダウン数の頻度データを解析した結果、ダウン数の分布が非常に裾野が広いことを示した。彼らの求めた確率分布関数は、ダウン数とともに確率が一気に減少する部分のあと、あるダウン数からは確率の減少が急激にゆるやかになり、かなりの割合まで確率が

<sup>†1</sup> 北里大学理学部

School of Science, Kitasato University

<sup>†2</sup> スタンダードアンドプアーズ

Standard & Poor's

\*1 サブプライム問題の影響が巨大化した理由は別であり、サブプライムローンを組み込んだABSのトランシェ（優先劣後構造を持つ証券）を別のCDOに組み込むといった入れ子構造が複雑に絡み合い、マーケット参加者が疑心暗鬼になってしまったことである。

残る．それ以上のダウン数の事象については，観測時間の問題があり観測できていないが，この確率分布関数の振舞いは興味深く，サーバのダウンが独立にはおきず非常に強い依存性を持つことを示している．なぜなら，独立にサーバがダウンするならば，確率分布は二項分布やポアソン分布と類似のものになり，その裾野は非常に狭いものになるはずだからである．彼らは，この確率分布を記述するために，ほぼ独立にダウンしていると考えられる小規模な現象（確率  $1 - \alpha$ ）を記述する指数分布（パラメータ  $r_1$ ）と大規模な連鎖ダウン（確率  $\alpha$ ）を記述する指数分布（パラメータ  $r_2$ ）を用意し，それらを重ね合わせた二重指数分布（Bi-Exponential）モデルを導入した．サーバ  $N$  台に対するダウン数  $n$  の確率分布関数  $P_N(n)$  は，

$$P_N(n) = (1 - \alpha)f(r_1, n) + \alpha f(r_2, n),$$

$$f(r, n) = C(r) \cdot r^n, \quad C(r) = \frac{1 - r}{1 - r^{N+1}} \quad (1)$$

で与えられる．データから推定されたパラメータの値は  $\alpha$  が  $0.009 \sim 0.0012$ ， $r_1$  が  $0.32 \sim 0.4$ ， $r_2$  が  $0.95 \sim 0.98$  であり，大規模な連鎖の起こる確率  $\alpha$  は  $0.1\%$  から  $1\%$  以下でしかないが， $r_2$  が非常に 1 に近いので，その確率がダウン数  $n$  とともに非常にゆっくりとしか減衰しないことが分かる．そして，この確率分布をもとに，しかし，サーバのダウンは独立に起こるとして，ネットワークの設計指針について議論した．つまり，ダウン数を与えたときに，ネットワーク上のどのサーバがダウンするとするかはランダムに選ぶわけ<sup>\*1</sup>である．

一方，大規模な連鎖ダウンのそれぞれの事象は，ウイルスの影響，DDoS 攻撃といった個々の理由がある．しかし，3 種類の異なるネットワークでほぼ同じような確率分布が得られたことは，なんらかの普遍的な性質を反映した結果であるとも考えられる．そこで，本論文では，ネットワーク上の相関のある二項分布モデルを導入し，次数が逆べき則に従う一般化ランダムグラフ<sup>8)</sup> や Barabasi らが提案したスケールフリーネットワークモデル（以下，BA モデルと呼ぶ<sup>9)</sup>）での確率分布関数を求めた．それらの振舞いは，二重指数分布と非常に似ており，サーバのダウン数の確率分布が二重指数分布のように振舞う原因の 1 つがネットワークのスケールフリー性であり，ダウン相関が非常に強いことを示唆している．ここで提案する確率モデルを用いれば，サーバネットワークのリスク評価をより正確に行えると考える．

\*1 これは，サーバネットワークが完全グラフだと仮定することと同じ．

まず，2 章では，ネットワーク上の相関二項分布の構成方法を説明する．条件付き確率を用いて，ネットワーク上のある確率変数から逐次，期待値，相関を固定しながら同時分布関数を構成する．この手法で同時分布が一意に構成できる場合とできない場合について詳細な解説を行う．3 章では，さまざまなネットワーク上の相関二項分布の振舞いについて説明する．1 次元ネットワークの場合は，確率分布関数の展開公式とその連続極限を厳密に求めることができる．ケーリーツリーや一般化ランダムグラフの場合は，確率分布を再帰的な関係式を用いて求めることが可能であり，その振舞いを調べることができた．次の 4 章では，より現実的なスケールフリーネットワークとして BA モデル上での相関のある二項分布モデルを扱う．そして，平均相関，各変数の期待値を同じにしたとき，一般化ランダムグラフ上の相関二項分布とほぼ同じような振舞いをするを示す．また，二重指数分布モデルとも類似性を持つことも示す．この事実から，サーバのダウンの相関が非常に大きいことが分かる．最後の 5 章では，得られた結果のまとめと，今後の課題について述べる．付録では，二次元正方形格子上の相関二項分布の構成方法と相関関数，確率分布関数について簡単に述べた．

## 2. 相関二項分布の構成方法

ネットワークは  $N$  個のノードからなり，1 から  $N$  までの整数の集合  $V = \{1, 2, \dots, N\} \equiv \{i\}_{i=1, \dots, N}$  で表すことにする．各ノード上に二値の確率変数  $X_i = 0, 1$  をおく．ノード 2 個の組からなる集合  $B = \{(i, j)\}$  を作り，その各要素  $(i, j)$  をリンクと呼ぶ．目標は各確率変数  $X_i$  の期待値  $p_i$  とリンク  $(i, j)$  で結ばれた変数  $X_i, X_j$  間に相関の相関  $\rho_{i,j}$  を与えたときに， $N$  個の変数全体での同時分布関数を構成することである．相関  $\rho_{i,j}$  とは，ピアソンの相関係数のことで，次の式で定義される．

$$\rho_{i,j} \equiv \frac{\langle X_i X_j \rangle - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}$$

ここで， $\langle A \rangle$  とは確率変数  $A$  の期待値を表す．また，同時分布関数  $P(\{x_i\}_{i=1, \dots, N})$  とは，各確率変数  $X_i$  が  $x_i$  をとる確率を表す．

$$P(\{x_i\}_{i=1, \dots, N}) \equiv \text{Prob.}(\{X_i = x_i\}_{i=1, \dots, N})$$

同時分布は規格化の制限から  $2^N - 1$  の自由度を持つ．一方，期待値の条件は  $N$  個，相関の条件はすべてのノードがリンクされた完全グラフの場合でも  ${}_N C_2$  個の条件なので，期待値と相関だけで同時分布が決定されるのは  $N = 2$  の場合のみである． $N > 2$  の場合は，同時分布の構成方法を具体的に指定する必要がある．そこで，確率変数を逐次つなげていく

過程で用いる条件付き確率の関数形を指定し、そのパラメータを期待値と相関から定める方針をとることにする。

2.1 同時分布関数の構成方法

$N$  個の確率変数  $\{X_i\}_{i=1,\dots,N}$  の従う同時分布関数  $p(\{x_i\}_{i=1,\dots,N})$  を以下の手順で構成する。

- (1) ノードの集合  $V$  から重複を許さずに  $N$  個のノードを選び順序を決める。それを、 $\{t_i\}_{i=1,\dots,N} = \{t_1, \dots, t_N\}$  で表す。最初のノードはランダムに選ぶ。それ以降のノードは、それまでに選ばれたノードのいずれかとリンクされているノードを選ぶものとする。たとえば  $j$  番目のノード  $t_j$  を選ぶとき、それまでのノード  $t_1, \dots, t_{j-1}$  のいずれかとリンクで結ばれているノードを選ぶ。
- (2) 前の手順で定めた順序に従い、同時分布を構成する。まず、ノード  $t_1$  上の確率変数  $X_{t_1}$  の期待値は  $p_{t_1}$  なので、 $X_{t_1}$  の従う確率分布関数は

$$P(x_{t_1}) = p_{t_1}^{x_{t_1}} \cdot q_{t_1}^{1-x_{t_1}}$$

ここで、 $q_{t_i} \equiv 1 - p_{t_i}$  は、確率変数  $X_{t_i}$  がゼロをとる確率を表す。

- (3) 次のノード  $t_2$  上の確率変数  $X_{t_2}$  は、期待値は  $p_{t_2}$  で、変数  $X_{t_1}$  とリンクで結ばれている。そこで、 $X_{t_1}$  への依存性を考えて、次の条件付き確率に従ってランダムに振る舞うとする。

$$p(x_{t_2} = 1|x_{t_1}) \equiv \text{Prob.}(X_{t_2} = x_{t_2} = 1|X_{t_1} = x_{t_1}) = \alpha_2 + \beta_2 \cdot x_{t_1} \quad (2)$$

$$p(x_{t_2} = 0|x_{t_1}) = 1 - p(x_{t_2} = 1|x_{t_1}) \quad (3)$$

確率変数  $X_{t_1}, X_{t_2}$  の従う同時分布  $p(x_{t_1}, x_{t_2})$  は次の式で構成する。

$$p(x_{t_1}, x_{t_2}) = p(x_{t_1}) \cdot p(x_{t_2}|x_{t_1})$$

$N = 2$  ならここで終了。

- (4) 3 番目のノード  $t_3$  は、ノード  $t_1, t_2$  の少なくとも一方とはリンクされている。一方のみとリンクされている場合、そのノードを  $t_{3,1}$  で表すと、2 番目のノードと同様な条件付き確率で  $x_{t_{3,1}}$  に依存する条件付き確率で同時分布を構成する。

$$p(x_{t_3} = 1|x_{t_{3,1}}) \equiv \text{Prob.}(X_{t_3} = x_{t_3} = 1|X_{t_{3,1}} = x_{t_{3,1}}) = \alpha_3 + \beta_{3,1} \cdot x_{t_{3,1}} \quad (4)$$

$$p(x_{t_3} = 0|x_{t_{3,1}}) = 1 - p(x_{t_3} = 1|x_{t_{3,1}}) \quad (5)$$

同時分布関数  $p(x_{t_1}, x_{t_2}, x_{t_3})$  は、条件付き確率  $p(x_{t_3}|x_{t_{3,1}})$  と  $p(x_{t_1}, x_{t_2})$  との積により構成する。

$$p(x_{t_1}, x_{t_2}, x_{t_3}) = p(x_{t_1}, x_{t_2}) \cdot p(x_{t_3}|x_{t_{3,1}})$$

ノード  $t_3$  がノード  $t_1, t_2$  の双方にリンクされている場合は、これらのノードを  $t_{3,1}, t_{3,2}$  と表す。そして、次の条件付き確率を用いる。

$$p(x_{t_3} = 1|x_{t_{3,1}}, x_{t_{3,2}}) = \alpha_3 + \beta_{3,1} \cdot x_{t_{3,1}} + \beta_{3,2} \cdot x_{t_{3,2}} \quad (6)$$

$$p(x_{t_3} = 0|x_{t_{3,1}}) = 1 - p(x_{t_3} = 1|x_{t_{3,1}}) \quad (7)$$

同時分布関数  $p(x_{t_1}, x_{t_2}, x_{t_3})$  は、条件付き確率に  $p(x_{t_3}|x_{t_{3,1}}, x_{t_{3,2}})$  を用いて構成する。 $N = 3$  ならここで終了。

- (5)  $k \geq 4$  番目のノード  $t_k$  の場合を考える。このノードより順序が前で、かつこのノードにリンクされているノードを  $\{t_{k,i}\}_{i=1,\dots,n_k}$  で表す。そして次の条件付き確率を用意する。

$$p(x_{t_k} = 1|\{x_{t_{k,i}}\}_{i=1,\dots,n_k}) = \alpha_k + \sum_{i=1}^{n_k} \beta_{k,i} \cdot x_{t_{k,i}}$$

$$p(x_{t_k} = 0|\{x_{t_{k,i}}\}_{i=1,\dots,n_k}) = 1 - p(x_{t_k} = 1|\{x_{t_{k,i}}\}_{i=1,\dots,n_k}) \quad (8)$$

同時分布関数  $p(\{x_{t_i}\}_{i=1,\dots,k})$  は、 $p(\{x_{t_i}\}_{i=1,\dots,k-1})$  と  $p(x_{t_k}|\{x_{t_{k,i}}\}_{i=1,\dots,n_k})$  の積により構成される。

$$p(\{x_{t_i}\}_{i=1,\dots,k}) = p(\{x_{t_i}\}_{i=1,\dots,k-1}) \cdot p(x_{t_k}|\{x_{t_{k,i}}\}_{i=1,\dots,n_k}) \quad (9)$$

- (6) 操作 (5) を  $k = N$  まで繰り返す。

ノード  $t_k$  上の確率変数  $X_{t_k}$  を接続する際に用いる条件付確率分布  $p(x_{t_k}|\{x_{t_{k,i}}\}_{i=1,\dots,n_k})$  が  $t_k$  より順序が前の確率変数  $\{x_{t_{k,i}}\}_{i=1,\dots,n_k}$  の線形結合に依存していることが重要である。ベイジアンネットでは、条件付き確率の対数が依存する変数の多項式で表現される<sup>10)</sup>。一方、条件付き確率が変数の線形結合に依存する場合は以下で見ると期待値や相関をコントロールしながらパラメータ  $\{\alpha_k, \beta_{k,i}\}_{i=1,\dots,n_k}$  を決定できる場合があること、また、相関関数がノード間の距離とともに指数関数的に減衰するという性質を持つ。特に後者は相関二項分布の持つべき性質だと考えられること、前者は構成された確率分布をコントロールしやすいことから、この関数形の条件付き確率を採用した。

条件付き確率のパラメータの決定方法にうつる。まず、確率変数の期待値  $\langle X_{t_k} \rangle = p_{t_k}$  を実現するには、次の関係式を満たす必要がある。

$$p_{t_k} = \alpha_k + \sum_{i=1}^{n_k} \beta_{k,i} p_{t_{k,i}} \quad (10)$$

この条件は、

$$\langle X_{t_k} \rangle = \langle \alpha_k + \sum_{i=1}^{n_k} \beta_{k,i} X_{t_{k,i}} \rangle$$

から導かれる．問題はリンクされている変数間の相関を実現することである．変数  $X_{t_k}$  と  $X_{t_{k,i}}$  間の相関は  $\rho_{t_k, t_{k,i}}$  なので，その積の期待値は

$$\langle X_{t_k} X_{t_{k,i}} \rangle = \sigma_{t_k} \sigma_{t_{k,i}} \rho_{t_k, t_{k,i}} + p_{t_k} \cdot p_{t_{k,i}}$$

となる．ここで， $\sigma_i$  で変数  $X_i$  の標準偏差を表し ( $\sigma_i^2 = p_i \cdot q_i$ )，また， $\tilde{\rho}_{i,j} = \rho_{i,j} \cdot \sigma_i \cdot \sigma_j$  を定義すると，

$$\langle X_{t_k} X_{t_{k,i}} \rangle = \tilde{\rho}_{t_k, t_{k,i}} + p_{t_k} \cdot p_{t_{k,i}} \quad (11)$$

とまとめることができる．一方， $\langle X_i X_j \rangle$  は， $p_i \cdot \langle X_j | X_i = 1 \rangle$  と分解できるので式 (11) の左辺は次のように変形できる．

$$p_{t_{k,i}} \cdot \left( \alpha_k + \sum_{j=1, j \neq i}^{n_k} \beta_{k,j} \langle X_{t_{k,j}} | X_{t_{k,i}} = 1 \rangle + \beta_{t_{k,i}} \right).$$

変数  $X_{t_{k,i}}$ ， $X_{t_{k,j}}$  間の相関  $\rho_{t_{k,i}, t_{k,j}}$  を用いてさらに変形すると，次の式になる．

$$\tilde{\rho}_{t_k, t_{k,i}} = \sum_{j=1, j \neq i}^{n_k} \tilde{\rho}_{t_{k,i}, t_{k,j}} \beta_{k,j} + \sigma_{t_{k,i}}^2 \cdot \beta_{k,i} \quad (12)$$

$\rho_{i,i} \equiv 1$  と定義すると， $\tilde{\rho}_{i,i} = \sigma_i^2$  となり， $\beta_{k,j}$  の満たすべき式は

$$\tilde{\rho}_{t_k, t_{k,i}} = \sum_{j=1}^{n_k} \tilde{\rho}_{t_{k,i}, t_{k,j}} \beta_{k,j} \quad (13)$$

とまとめることができる．これを解いて，得られた係数  $\beta_{k,i}$  を式 (10) に代入すると，係数  $\alpha_k$  も求めることができる．

この一連のプロセスを実行する際に問題となるのが，相関  $\rho_{t_{k,i}, t_{k,j}}$  の評価である．これが最初から分かっている場合は，式 (13) を解くことにより実行可能である．たとえば，ツリー型のループを含まないネットワークの場合，ノードを接続する場所は 1 か所なので， $n_k = 1$  となり，解くべき式は

$$\tilde{\rho}_{t_k, t_{k,1}} = \sigma_{k,1}^2 \beta_{k,1} \quad (14)$$

となる．係数は

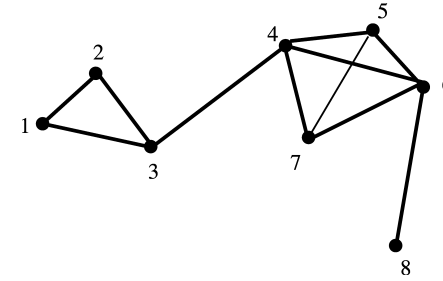


図 1 ネットワークの例．ノードを  $\bullet$ ，リンクを実線で描いた．ノードの集合は  $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ ，リンクの集合は  $B = \{(1, 2), (1, 3), (2, 3), (3, 4), (4, 5), (4, 6), (4, 7), (5, 6), (6, 7), (6, 8)\}$   
 Fig. 1 Example of a network.  $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$  and  $B = \{(1, 2), (1, 3), (2, 3), (3, 4), (4, 5), (4, 6), (4, 7), (5, 6), (6, 7), (6, 8)\}$ .

$$\beta_{k,1} = \frac{\tilde{\rho}_{t_k, t_{k,1}}}{\sigma_{t_{k,1}}^2}, \quad \alpha_k = p_{t_k} - \beta_{k,1} p_{k,1} \quad (15)$$

と求めることができる．特に，すべての変数の期待値が等しい場合 ( $p_i = p$  for all  $i$ )，

$$\beta_{k,1} = \rho_{t_k, t_{k,1}}, \quad \alpha_k = p(1 - \rho_{t_k, t_{k,1}})$$

となる．また，ノードが接続されるノードの集合が，それらだけで完全グラフを構成する場合は，その完全グラフのリンク上で相関係数  $\rho_{t_i, t_j}$  がすべて決まっているので問題ない．

図 1 で示されたネットワークを考えてみよう．このネットワーク上の確率変数に対して，ノード番号と同じ順序で同時分布関数を構成する．まず，ノード 1 は出発点なので，期待値から  $p(x_1)$  を構成できる．ノード 2 がつながっている順序が前のノードはノード 1 のみなので， $p(x_2|x_1)$  は，式 (15) から計算できる．

$$p(x_2|x_1) = \alpha_2 + \beta_{2,1} \cdot x_1, \quad \alpha_2 = p_2 - \beta_{2,1} p_1, \quad \beta_{2,1} = \frac{\tilde{\rho}_{1,2}}{\sigma_1^2}$$

$p(x_1)$  との積を計算して  $p(x_1, x_2)$  を構成できる．次にノード 3 だが，これと接続する順序が前のノードは 1 と 2 だが，1 と 2 はもともとリンクされていて相関  $\rho_{1,2}$  が与えられているので， $\beta_{3,1}, \beta_{3,2}$  は式 (10)，(13) を解いて求めることができる．得られた  $p(x_3|x_1, x_2)$  を用いて， $p(x_1, x_2, x_3)$  も構成できる．ノード 4, 5, 6 までは同様なプロセスを繰り返し，ノード 7 についても依存するノード 4, 5, 6 がそれらで完全グラフを形作っているので問題なくパラメータ  $\alpha_7, \beta_{7,i}$  をもとめ，同時分布  $p(\{x_i\}_{i=1, \dots, 7})$  を構成できる．ノード 8 も

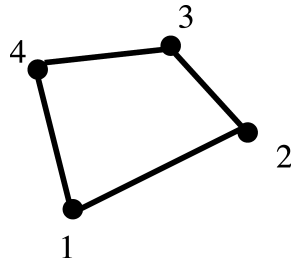


図2 ネットワークの例.  $N = 4$  で, ノードを  $\bullet$ , リンクを実線で描いた. ノードの集合は  $V = \{1, 2, 3, 4\}$ , リンクの集合は  $B = \{(1, 2), (1, 4), (2, 3), (3, 4)\}$

Fig. 2 Example of a network.  $V = \{1, 2, 3, 4\}$  and  $B = \{(1, 2), (1, 4), (2, 3), (3, 4)\}$ .

ノード 7 のみに依存するので, 解 (15) を用いて同時分布  $p(\{x_i\}_{i=1, \dots, 8})$  を求めることができる.

では, 変数を接続していく過程で, 被接続ノードが完全グラフを構成していない場合はどうするのか? この場合は, 被接続ノード間でリンクがない部分すべての相関係数をそのつど求め, その相関係数で変数が関連している完全グラフの問題に帰着させ, 上記の手順を行う必要がある. 図2の場合に同時分布関数を構成してみよう. ノードの番号順に変数を接続していくものとする. ノード1から3までは, 解(15)を用いて同時分布関数  $p(x_1, x_2, x_3)$  を構成できる. 問題は変数  $X_4$  を接続するとき起こる. ノード4はノード1とノード3にリンクしているが, ノード1とノード3はリンクされていず, ノード2を経由して間接的に依存している. 条件付き確率  $p(x_4|x_1, x_3)$

$$p(x_4 = 1|x_1, x_3) = \alpha_4 + \beta_{4,1}x_1 + \beta_{4,2}x_3$$

の係数  $\alpha_4, \beta_{4,1}, \beta_{4,2}$  を決定するには,  $X_1$  と  $X_3$  の相関係数の値が必要になる.  $p(x_1, x_2, x_3)$  を用いて計算すると,

$$\rho_{1,3} = \rho_{1,2} \cdot \rho_{2,3}$$

となることが分かる. よって, 係数の満たすべき条件は次のようになる.

$$\begin{aligned} p_4 &= \alpha_4 + \beta_{4,1} \cdot p_1 + \beta_{4,2} \cdot p_3 \\ \tilde{\rho}_{4,1} &= \tilde{\rho}_{1,1} \cdot \beta_{4,1} + \tilde{\rho}_{1,3} \cdot \beta_{4,1} \\ \tilde{\rho}_{4,2} &= \tilde{\rho}_{3,1} \cdot \beta_{4,1} + \tilde{\rho}_{3,3} \cdot \beta_{4,2} \end{aligned} \tag{16}$$

特に,  $p_1 = p_2 = p_3 = p_4 = p, \rho_{1,2} = \rho_{2,3} = \rho_{3,4} = \rho_{1,4} = \rho$  の場合は,

$$\alpha_4 = p \frac{(1-\rho)^2}{1+\rho^2} \quad \beta_{4,1} = \beta_{4,2} = \frac{\rho}{1+\rho^2} \tag{17}$$

と求めることができる.

以上の説明から, ツリー型のネットワーク, 図1のように部分的に完全グラフを持つツリー型のネットワーク, そしてループを持つが単純な構造のネットワークの場合には与えられた期待値, 相関係数を再現するようにノードの順序に従って対応する条件付き確率を逐次決定し, 同時分布関数を構成することが可能である. 一方, 複雑な構造のネットワークの場合は, ノードを逐次接続して追加された変数と, それまでの変数のすべての相関が計算できる場合, 次のノードの接続でのパラメータ, 条件付き確率を決定できる. つまり, 毎回の接続において相関係数をすべてのペアで計算でき, 完全グラフだと見なせる場合は, 次の接続も可能となる. しかし, それは必ずしも容易とは限らない. また, 以下で見るように, ループを含むネットワークの場合は, 同時分布関数を一意に構成できないという別の困難もある.

### 2.2 同時分布関数の順序依存性

前節では, ノードに順序  $\{t_i\}_{i=1, \dots, N}$  をつけ, それに従って条件付き確率をかけていくことで同時分布関数を構成した. では, 順序とはどのような意味を持つのか? 確率変数がサーバのダウンや債務者のデフォルトを表すのなら, ノードに順番をつけるのは, ダウンやデフォルトの順番の指定に対応していると考えられる. 実際のダウンやデフォルトでは, どの順序で起こるのかは分からないため, もし同時分布関数が順序に依存するなら, あらゆる順序についての平均操作が必要となる. しかし, 同時分布関数が順序に依存しないなら, 平均操作は不必要で, ある1つの順序での同時分布関数を考えればよいことになる. この節では, 同時分布関数が, 最初の確率変数の順序には依存しないための十分条件を述べる.

まず, ネットワークがツリー構造を持ち, ループがいない場合, 同時分布関数は順序には依存しない. このことは, 次の事実から自明である. まず, ツリーの場合, あるノード  $t_1$  を出発点にして,  $t_2, t_3$  とノードを接続していく. その際, 接続するノード  $t_k$  は, 被接続ノードとして親にあたる1個  $t_{k,1}$  しか持たないため,  $p(x_{t_k}|x_{t_{k,1}})$  をかけることで同時分布を構成することになる. たとえば,  $k=2$  の場合,  $t_{2,1}$  は出発したノード  $t_1$  のことで, こままでの同時分布関数は

$$p(x_{t_1}, x_{t_2}) = p(x_{t_1}) \cdot p(x_{t_2}|x_{t_1})$$

となる. この同時分布は  $\langle X_{t_1} \rangle = p_{t_1}, \langle X_{t_2} \rangle = p_{t_2}$  という期待値と,  $X_{t_1}$  と  $X_{t_2}$  の相関係数が  $\rho_{t_1, t_2}$  を実現するように構成されている. ここで重要なのは, 本章の最初に説明したとおり, これらの期待値と相関係数の条件を満たす同時分布は  $N=2$  の場合は一意に定まることである. そのため,  $t_1, t_2$  の順序を入れ換えて,  $t_2, t_1$  の順番で同時分布を構成し

ても同じものが得られる．よって次の等式が成り立つ．

$$p(x_{t_1}) \cdot p(x_{t_2}|x_{t_1}) = p(x_{t_2}) \cdot p(x_{t_1}|x_{t_2})$$

同様に，ノード  $t_3$  がノード  $t_1$  に接続されるなら，同時分布  $p(x_{t_1}, x_{t_2}, x_{t_3})$  は， $p(x_{t_1}) \cdot p(x_{t_2}|x_{t_1}) \cdot p(x_{t_3}|x_{t_1})$  と構成されるが， $p(x_{t_1}) \cdot p(x_{t_3}|x_{t_1})$  に着目すると， $x_{t_1}, x_{t_2}$  の同時分布と同様に，期待値，相関係数の条件から一意に定まっており， $t_1, t_3$  の順序を入れ換えても同じものが得られる．

$$p(x_{t_1}) \cdot p(x_{t_3}|x_{t_1}) = p(x_{t_3}) \cdot p(x_{t_1}|x_{t_3})$$

よって， $p(x_{t_1}, x_{t_2}, x_{t_3})$  に関して  $t_1, t_2, t_3$  の順序を入れ換えても同時分布は変化せず，次の等式が成り立つことになる．

$$\begin{aligned} p(x_{t_1}) \cdot p(x_{t_2}|x_{t_1}) \cdot p(x_{t_3}|x_{t_1}) &= p(x_{t_2}) \cdot p(x_{t_1}|x_{t_2}) \cdot p(x_{t_3}|x_{t_1}) \\ &= p(x_{t_3}) \cdot p(x_{t_1}|x_{t_3}) \cdot p(x_{t_2}|x_{t_1}) \end{aligned} \quad (18)$$

このように，同時分布は構成する順番には依存しない． $t_4, t_5$  と次々と変数を追加しても状況は同じで，出発点をツリー上の任意のノードにし，任意の順序で接続しても，同時分布関数は変化しない．よってツリー上の相関のある二項分布はノードの順序に依存せず，期待値，相関係数から一意に定まることが分かった．

次に，図 1 のように，ネットワークはループを持つが，ループを含む最小部分が完全グラフとなっている場合を考える．結論から述べると，完全グラフの部分で，期待値と相関係数がすべて等しく  $p, \rho$  となっている場合は変数の順序に依存しない．理由は，完全グラフ上での同時分布はベータ二項分布の同時分布となるからである．ベータ二項分布は二項分布をベータ関数で重ね合わせたもので，その同時分布関数は 1 や 0 をとる変数の個数にしか依存しないので，変数を交換しても変化しないからである．よって，完全グラフの部分で変数の順序に依存しないので，ネットワーク全体でも同時分布はノードの順序に依存しない．一方，期待値，相関係数がノード，リンクによって異なる場合は，同時分布関数はノードの順序に依存する．

では，ネットワークにループが存在するが，そのループを含む最小の部分グラフが完全グラフとならない場合を考える．たとえば，図 2 の場合である．この場合，同時分布関数は順序に依存する．ノードの番号順に同時分布を構成した場合 ( $t_1 = 1, t_2 = 2, t_3 = 3, t_4 = 4$ ) と， $t_1 = 1, t_2 = 2, t_3 = 4, t_4 = 3$  と， $t_3, t_4$  を入れ換えた場合では，同時分布が異なる．このように，ループが存在する場合は，ノードの順序に同時分布は依存してしまう．最初の順序ではサーバ 3 がダウンしてからサーバ 4 がダウンする場合，サーバ 1 の状態に依存するが，2 番目に考えた順序では，サーバ 4 がダウンしてからサーバ 3 がダウンす

る場合は，サーバ 1 の状態には直接依存しないからである．こうした場合は，あらゆる順序を考え，順序に関する平均化の操作を行う必要が出てくる．問題は，平均化をどのように行うのが正しいのかである．単に，すべての順序が同じ確率で実現すると考えた単純平均なのか，それとも順序によって重みが異なるのか．これらの問題は現時点では未解決である．また，同時分布が順序に依存するとしても，その依存性が強いのか，弱いのか．同じ期待値，相関係数をあたえて，同じ関数形を持つ条件付き確率で同時分布を構成しているのだから，多少は順序に依存しても，ノードの数が十分大きくなれば自己平均化により，順序に依存しなくなる可能性もある．その場合は，適当な順序の場合に同時分布を構成して性質を調べれば十分となる．しかし，この問題の解答も現時点では分からない．

### 3. 確率分布の連続極限とネットワーク構造依存性

この章では，ループを持たないネットワーク上の相関のある二項分布の性質を調べてみる．特に，確率分布関数  $P_N(n) \equiv \text{Prob.}(\sum_{i=1}^N X_i = n)$  と次式で定義される確率密度関数  $p(x)$  がネットワーク構造にどのように依存しているのかに注目する．

$$p(x) \equiv \lim_{N \rightarrow \infty} N \cdot P_N(n = N \cdot x) \quad (19)$$

以下，期待値はすべての確率変数で  $p$ ，相関係数は  $\rho$  とする．もちろん，リンクで直接結び付いた確率変数の相関係数が  $\rho$  という意味であって，直接結び付いていない変数間の相関は一般に  $\rho$  とは異なっている．

#### 3.1 1次元格子上の相関のある二項分布

ノードの集合は  $V = \{1, 2, \dots, N\}$ ，リンクの集合は  $B = \{(i, i+1)\}_{i=1, \dots, N-1}$  の 1次元格子を考える．1次元格子はループを持たないツリー構造のネットワークの一種なので，構成する同時分布はノードの順序に依存しない．そこで， $t_i = i$  とノードの番号の順序を用いることとする．

まず， $\langle X_1 \rangle = p$  より，

$$p(x_1) = p^{x_1} \cdot q^{1-x_1}$$

ここで， $q \equiv 1-p$  と定義する． $i \geq 2$  以降の  $X_i$  は， $X_{i-1}$  に次の条件付確率で連結していればよい．

$$p(x_i|x_{i-1}) = (p(1-\rho) + \rho \cdot x_{i-1})^{x_i} \cdot (q(1-\rho) + \rho \cdot (1-x_{i-1}))^{1-x_i} \quad (20)$$

よって，同時分布関数は次の式で書くことができる．

$$p(\{x_i\}_{i=1,\dots,N}) = p(x_1) \cdot \prod_{i=2}^N p(x_i|x_{i-1})$$

相関係数  $\rho_{i,j}$  を計算する．同時分布関数から，条件付き確率  $p(x_j|x_i)$  を計算すると

$$p(x_j|x_i) = (p(1 - \rho^{|j-i|}) + \rho^{|j-i|} \cdot x_i)^{x_j} \cdot (q(1 - \rho^{|j-i|}) + \rho^{|j-i|} \cdot (1 - x_i))^{1-x_j} \quad (21)$$

となり， $\rho_{i,j} = \rho^{|j-i|}$  と振る舞う．相関係数はノード間の距離にしか依存しないので

$$\rho_k \equiv \rho_{i,i+k} = \rho^k$$

と書くことにする．相関関数はノードの距離の増加とともに指数関数的に減少することが分かる．

次に確率分布関数について調べる．ここで条件付き確率分布関数を次の式で定義する．

$$P_N(n|x_N) = \text{Prob.} \left( \sum_{i=1}^N X_i = n | X_N = x_N \right)$$

後者は， $X_N$  の値を  $x_N$  決めた条件での確率分布関数である．もちろん， $x_N$  について和をとれば，無条件の確率分布関数  $P_N(n)$  となる．同時分布関数の構成方法から  $P_N(n|x_N)$  について次の再帰的な関係式を導ける．

$$P_{N+1}(n|x_{N+1}) = \sum_{x_N=0}^1 P_N(n-1|x_N)p(x_{N+1}|x_N)x_{N+1} + \sum_{x_N=0}^1 P_N(n|x_N)p(x_{N+1}|x_N)(1-x_{N+1}). \quad (22)$$

この関係式を用いると， $N = 10^4$  程度の確率分布を数値的に計算するのは容易である．

$P_N(N)$  が期待値  $p$ ，相関  $\rho$  の変化にともなってどう変化するのかを調べたい．しかし， $P_N(n)$  はその定義式から  $N$  に依存する．つまり，同じ  $p, \rho$  でも， $N$  の変化とともに  $P_N(n)$  は変わってしまう．特に，相関関数  $\rho_i$  は，ノード間の距離とともに急激に減少するため， $p, \rho$  を固定して  $N$  を大きくすると，隣接ノード間の相関  $\rho$  の効果はなくなり， $P_N(n)$  は二項分布に漸近する．この状況を見るために，次の平均相関  $\rho_{avg.}$  を導入する．

$$\rho_{avg.} \equiv \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} \rho_{i,j}$$

相関関数が求まっているので，次のような結果が得られる．

$$\rho_{avg.} = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \rho_i \cdot (N-i) = \frac{(N-1)\rho - N\rho^2 + \rho^{N+1}}{\frac{1}{2}N(N-1)(1-\rho^2)} \quad (23)$$

$\rho_{avg.}$  は， $\rho$  を固定して  $N$  を無限大にすると， $\rho_{avg.} = O(\frac{1}{N})$  なので，ゼロになる．平均相関がゼロになるということは，確率分布が二項分布に漸近的に近づくことを意味する．相関の確率分布関数への影響を調べるには， $N$  を大きくするとともに  $\rho$  を 1 に近づける極限操作を実行しなければならない<sup>13)</sup>．そこで， $\rho = 1 - \frac{\alpha}{N}$  という  $N$  依存性で  $\rho$  を 1 に近づけながら  $N$  の無限大の極限をとると，平均相関は

$$\rho_{avg.} = \frac{\alpha - 1 + e^{-\alpha}}{\frac{1}{2}\alpha^2}$$

となり，ゼロにならず有限の値をとることになる．

統計物理学におけるスピンモデルの低温展開の手法を応用して，確率分布関数に対する展開公式を導く． $\rho$  が 1 に近いと仮定し， $1 - \rho$  の多項式で展開する手法である．低温展開とのアナロジは，低温ではスピン間の相関が強くなることから由来する．結果を書くと

$$\begin{aligned} P_N(0) &= qq_+^{N-1}, \quad P_N(N) = pp_+^{N-1} \\ P_N(n) &= \sum_{l=0}^{\min(n-1, N-n-1)} (1-\rho)^{2l+1} 2 \cdot p^{l+1} q^{l+1} \cdot {}_{N-n-1}C_{ln-1} C_l p_+^{n-(l+1)} q_+^{(N-n)-(l+1)} \\ &+ \sum_{l=0}^{\min(n-1, N-n-2)} (1-\rho)^{2l+2} \cdot p^{l+1} q^{l+2} \cdot {}_{N-n-1}C_{l+1n-1} C_l p_+^{n-(l+1)} q_+^{(N-n)-(l+2)} \\ &+ \sum_{l=0}^{\min(n-2, N-n-1)} (1-\rho)^{2l+2} \cdot p^{l+2} q^{l+1} \cdot {}_{N-n-1}C_{ln-1} C_{l+1} p_+^{n-(l+2)} q_+^{(N-n)-(l+1)} \end{aligned}$$

となる．ここで， $l$  というのは，0 と 1 の間の境界の数を表す．また， $p_+ = p(1-\rho) + \rho, p_- = p(1-\rho), q_+ = q(1-\rho) + \rho, q_- = q(1-\rho)$  と定義した．特に， $N$  を固定したまま， $p$  を小さく， $\rho$  を 1 に近づけた場合，展開公式の第 1 項の最低次の項 ( $l = 1$ ) の寄与が最も大きく，確率分布は近似的に幾何分布として振る舞うことが分かる．

$$P_N(n) \propto \left( \frac{p_+}{q_+} \right)^n \quad (24)$$

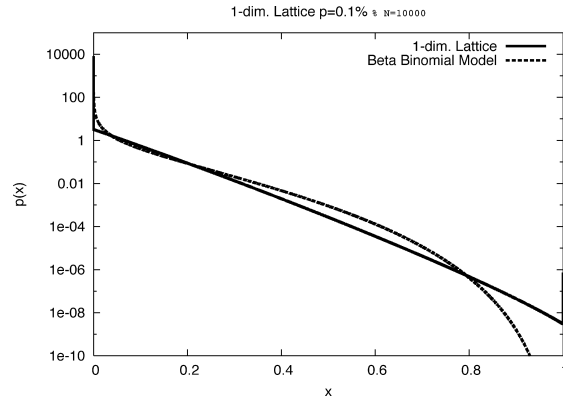


図 3 1次元格子上の相関のある二項分布の確率密度関数を片対数プロットした。パラメータは  $N = 10000$ ,  $p = 1\%$ ,  $\rho_{avg.} = 10\%$  である。比較のため、同じパラメータでベータ二項分布の確率密度関数も点線でプロットした  
Fig.3 Semi-log plot of  $p(x)$  of the one dimensional model. We set  $p = 1\%$ ,  $\rho_{avg.} = 10\%$  and  $N = 10^4$ . We also plot the probability density of the beta-binomial distribution.

また、展開式で連続極限をとると、次の確率密度関数  $p(x)$  に対する展開式となる<sup>11),12)</sup>  
\*1 .

$$p(x) = \alpha \sum_{l=0}^{l+1} p^{l+1} q^{l+1} e^{-\alpha p} e^{-\alpha(1-2p)x} \cdot \left( \frac{2}{(l!)^2} (1-x)^l x^l + \frac{1}{l!(l+1)!} (q(1-x))^{l+1} x^l + p(1-x)^l x^{l+1} \right) + qe^{-\alpha p} \delta(x) + pe^{-\alpha q} \delta(1-x) \quad (25)$$

式 (22) を用いて確率密度関数  $p(x)$  を調べる。図 3 に、1次元格子上の相関のある二項分布の  $p(x)$  を片対数プロットした。 $N = 10^4$  の場合で、 $p = 1\%$ 、平均相関を  $\rho_{avg.} = 10\%$  にセットした場合の結果である。また、同じパラメータでベータ二項分布の確率密度関数もプロットしている。これは、完全グラフの場合の相関のある二項分布に対応する。まず、1次元格子上の確率密度関数は  $x = 0$  と  $x = 1$  にデルタ関数型のピークを持つことが分かる。これは、式 (25) の最後の項に対応する。また、 $0 < x < 1$  では確率密度関数  $p(x)$  は、ほぼ直線で振る舞うこともグラフから読み取ることができる。つまり、指数分布となっている。

以上のことから、1次元格子上の相関二項分布は、稀な ( $p$  が小さい) 相関が強い  $\rho \rightarrow 1$

\*1 文献 11) では、変形ベッセル関数を用いた別の表式が導出されている。

場合に指数分布のように振る舞う。また、両端に、デルタ関数型のピークも持つことが分かった。サーバネットワークのダウン数の確率分布で観測されたような二重指数分布的な振舞いは見られなかった。もちろん、サーバネットワークはインターネットの一部であり、インターネットはスモールワールド性、スケールフリー性を持つような複雑ネットワークである。そのどちらも持たない1次元格子上のモデルで二重指数分布が再現できなくても不思議ではない。そこで、次のサブセクションでは、ケーリーツリーの場合の相関のある二項分布を調べてみることにする。

### 3.2 ケーリーツリー上の相関のある二項分布

ケーリーツリーや一般化ランダムグラフと呼ばれるネットワークは、1つのノードを起点(根, ルート)として、そこからリンク(枝)が生え、リンクの先のノードを起点として、またリンクが生えといったプロセスを繰り返すことで生成される。ケーリーツリーの場合は、ノードから生える枝の数が、最初は  $z$  本で、次のノードからは  $z-1$  本と固定されている場合である。端のノード以外は、すべて次数  $z$  のネットワークとなる。一般化ランダムグラフは、ノードから生える枝の数  $z$  が確率変数の場合に生成されるネットワークである。 $z$  の従う確率分布は、インターネットのモデルを考える際には、そのスケールフリー性を考慮し、巾乗則の確率分布を用いる。まず、ケーリーツリーの場合から説明を行うが、 $z$  が確率変数になっているかなっていないかの差異しかないので、ほとんどの結論が一般化ランダムグラフの場合に適用できることを注意する。

ケーリーツリーの構造を詳しく指定し、ノード上の確率変数をラベルする(図4を参照)。ルート上の確率変数を  $X$  と書く。ルートはルートからの距離がゼロなので、レベルゼロであるとする。次に、ルートから距離1(レベル1)のところから  $z$  個のノードが存在する。それらを  $i_1 = 1, 2, \dots, z$  で指定し、対応する確率変数を  $X_{i_1}$  で表す。レベル1のノード  $i_1$  から  $z-1$  個の枝が生え、その先端の  $z-1$  個のノードを  $(i_1, i_2)$ 、対応する確率変数を  $X_{i_1, i_2}$  とラベルする。ルートから距離3のレベル3のノードも同様に、3つの整数の組  $(i_1, i_2, i_3)$ 、対応する確率変数を  $X_{i_1, i_2, i_3}$  で表す。一般に、レベル  $l$  のノードを  $l$  個の整数の組  $(i_j)_{j=1, \dots, l}$  で表し、対応する確率変数もそれを添字としてラベルする。

$$\text{ノード} : (i_1, \dots, i_l) \rightarrow X_{i_1, \dots, i_l}$$

ルートからの距離の最大値を最大レベルとして  $L$  で表すと、ノードの総数  $N_L$  は、

$$N_L = 1 + \sum_{l=1}^L z \cdot (z-1)^l = \sum_{l=0}^L B(l)$$



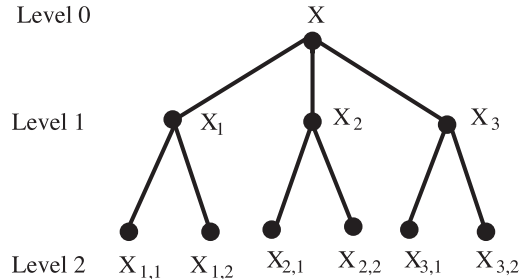


図4  $z = 3, L = 2$  のケーリーツリー. ノードの総数  $N_2 = 10$ .  
Fig. 4 Picture of Cayley Tree,  $z = 3, L = 2. N_2 = 10.$

ここで、レベル  $l$  のノードの数を  $B(l)$  で表すことにした。

ケーリーツリー上の相関のある二項分布は、ケーリーツリーにはループが存在しないため、順序に依存せずに同時分布関数を構成できる。出発点をルートのノードに選び、子ノードの条件付き確率を順次掛け合わせていけばよい。ただし、条件付き確率は、この章の最初に述べたように、すべての変数の期待値は  $p$ 、直接つながっているノード上の変数間の相関は  $\rho$  である。つまり、任意のノード  $(i_1, \dots, i_l)$  について次の条件を満たす。

$$\langle X_{i_1, \dots, i_l} \rangle = p \quad 0 \leq l \leq L$$

$$\rho_{(i_1, \dots, i_l), (i_1, \dots, i_l, i_{l+1})} = \rho \quad 0 \leq l \leq L-1$$

まず、相関関数だが、1次元格子の場合と同様にノード間の距離  $d$  にのみ依存し  $\rho^d$  と指数関数的に減衰する。1つのノードを  $(i_1, \dots, i_l, i_{l+1}, \dots, i_{l+j})$  もう1つのノード  $(i_1, \dots, i_l, i'_{l+1}, \dots, i'_{l+k})$  とし、レベル  $l$  までは共通の祖先を持っている場合を考える。つまり、レベル  $l+1$  のノードは異なるとする ( $i_{l+1} \neq i'_{l+1}$ )。このとき、このふたつのノードに対応する確率変数間の相関係数は

$$\rho_{(i_1, \dots, i_l, i_{l+1}, \dots, i_{l+j}), (i_1, \dots, i_l, i'_{l+1}, \dots, i'_{l+k})} = \rho^{j+k}$$

となる。このことは、ツリーの場合、順序非依存制が成立することから自明である。共通の祖先から出発する順序を考え、考えている2つのノードまでの先に条件付き確率を掛け合わせて同時分布を作ると、それは1次元格子上の相関のある二項分布モデルと同じである。その後で、他のノードを追加しても、以前に連結されたノード上の確率変数には影響を与えないので、1次元格子の結果をそのまま用いることができる。

この相関関数の結果を用いると、平均相関  $\rho_{avg.}$  は、次のように計算することができる。

$$\rho_{avg.} = \frac{\sum_{l=0}^L \sum_{m=0}^L \rho_B(l, m) - N_L}{N_L(N_L - 1)} \quad (26)$$

ここで、 $\rho_B(l, m)$  は、レベル  $l$  に属するノードとレベル  $m$  に属するノードの相関係数の和で定義され、 $l \leq m$  の場合次のように評価できる。

$$\begin{aligned} \rho_B(0, m) &= B(m) \cdot \rho^m \\ \rho_B(l, m) &= B(l) \cdot ((z-1)^m \rho^{l+m} + \sum_{i=1}^{l-1} (z-2)(z-1)^{m-i-1} \rho^{l+m-2i} \\ &\quad + (z-1)^{m-l} \rho^{m-l}) \quad \text{for } l \leq m \\ \rho_B(l, m) &= \rho_B(m, l) \quad \text{for } l \geq m \end{aligned} \quad (27)$$

確率分布関数を求めるには1次元格子のときと同じように、再帰的な関係式を数値的に解くしかない。まず、レベル  $L-1$  の親のノードにある確率変数を  $x$  に固定したとき、レベル  $L$  の子ノードの確率分布は、

$$P_L(n|x) = {}_{z-1}C_n \cdot p(1|x)^n \cdot p(0|x)^{(z-1)-n}$$

となる。これらの条件付き確率分布関数を  $z-1$  個束ねて、レベル  $L-2$  の親ノード上の確率変数を固定したときの、その子孫(レベル  $L-1, L$ ) の条件付き確率分布関数は

$$P_{L-1}(n|x) = \prod_{i=1}^{z-1} \left[ \sum_{x_i} \sum_{n_i} P_L(n_i|x_i) \cdot p(x_i|x) \right]$$

となる。この操作を順次レベル0まで繰り返す。ただし、レベル0においては、束ねる確率分布は  $z$  個となるので、次の関係式を用いる必要がある。

$$P_0(n|x) = \prod_{i=1}^z \left[ \sum_{x_i} \sum_{n_i} P_1(n_i|x_i) \cdot p(x_i|x) \right]$$

以上の手順を用いて求めた確率密度関数をプロットしたものが図5である。 $z = 3 (L = 14)$ ,  $z = 6 (L = 6)$  で、期待値は0.1%とし、平均相関は5%、 $z = 3$  の場合は同10%の分布も載せている。

このグラフから、ダウン数が少ない事象の確率はダウン数とともに急速に減少するが、ある規模以上では確率の減少がゆるやかになることが分かる。その意味では、二重指数分布の振舞いに近い。また、相関を増やす ( $\rho_{avg.} = 5\% \rightarrow 10\%$ ) と、分布の大規模な連鎖部分の

### 31 ネットワーク上の相関のある二項分布

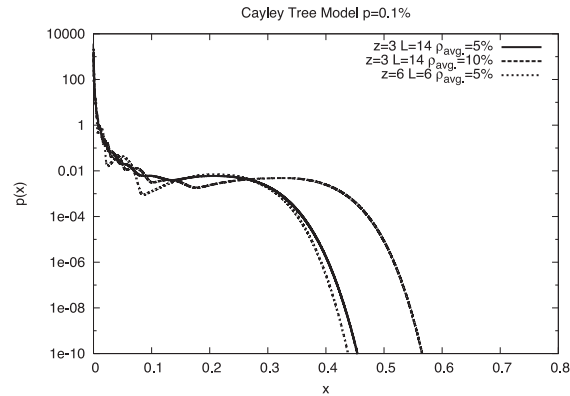


図5 ケーリーツリー上の相関のある二項分布の確率密度関数をプロットした。パラメータは  $z = 3, L = 14, p = 0.1\%, \rho_{avg.} = 5\%$ ,  $z = 3, L = 14, p = 0.1\%, \rho_{avg.} = 10\%$ , および  $z = 6, L = 6, p = 0.1\%, \rho_{avg.} = 5\%$  である

Fig. 5 Plot of  $p(x)$  of the cayley tree model. We set  $z = 3, L = 14, p = 0.1\%, \rho_{avg.} = 5\%$ ,  $z = 3, L = 14, p = 0.1\%, \rho_{avg.} = 10\%$ , and  $z = 6, L = 6, p = 0.1\%, \rho_{avg.} = 5\%$ .

確率は変化せずに、裾野だけがのび、その分、小規模な連鎖の確率が減少することも分かる。そのほか、確率分布が波打つ振舞いもグラフから読み取ることができる。特に、 $z = 6$  の場合の振動はかなり激しい。こうした振舞いは、サーバネットの連鎖ダウンの確率分布では確認されていない\*1。

以上の観察から、ケーリーツリー上の相関のある二項分布は、確率分布関数として興味深い性質をいろいろ持っているが、サーバネットの連鎖ダウンの確率モデルとしては問題もあることが分かる。そこで、次数  $z$  がスケールフリー性を持つネットワーク上での相関のある二項分布を導入し、その性質を調べてみよう。

#### 3.3 一般化ランダムグラフ上の相関のある二項分布

一般化ランダムグラフ<sup>8)</sup> というのは、ケーリーツリーにおいて固定されていた次数  $z$  が、確率変数となっている場合である。従う確率分布  $P_{degree}(z)$  として、インターネットのスケールフリー性を考慮し、次のモデルを考えることにする。

\*1 観測時間の問題から、確率分布を正確に推定できるほどのデータはないので、必ずしも断言はできない。実際、データのプロットはかなりギザギザとなっている。

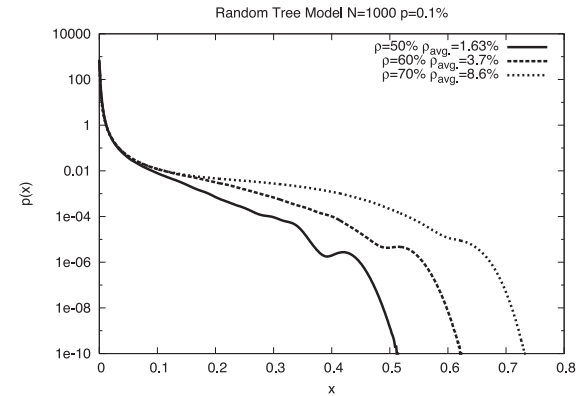


図6 一般化ランダムグラフ上の相関のある二項分布の確率密度関数をプロットした。共通のパラメータは  $p = 0.1\%$ ,  $N = 10^3$  とし、 $10^3$  個のネットワークサンプルで平均したものである。隣接ノード間の相関には、 $\rho = 50\%$ ,  $60\%$ ,  $70\%$  を用いている

Fig. 6 Plot of  $p(x)$  of the random tree model. We set  $p = 0.1\%$ ,  $N = 10^3$  in common. About  $\rho$ , we use  $\rho = 50\%, 60\%, 70\%$ .

$$P_{degree}(z) = \frac{C}{z^r} \quad \text{for } 3 \leq z \quad (28)$$

係数  $C$  は規格化の条件から定める。べきの指数  $r$  として、後に Barabási らの BA モデル<sup>9)</sup> 上の相関のある二項分布との比較を行うので  $r = 3$  としている。この確率法則に従い、ノードごとに乱数  $z$  を発生させ、ルートの場合は  $z$  本の枝を張り、それ以外のノードでは親からの枝を考慮して  $z - 1$  本の枝を張るという過程をノードの数が  $N$  となるまで繰り返す。以下、 $N = 10^3$  の場合に固定する。また、確率分布は、ノードの順序には依存しないがネットワークの構造に依存するので、ネットワークのサンプルを  $10^3$  個用意し、確率分布はネットワークサンプルによる平均も行う。

一般化ランダムグラフの場合、平均相関を計算する簡単な公式がないため、隣接ノード間の相関  $\rho$  をいろいろ変えたときの確率分布を調べた。図6に、 $\rho = 50\%$ ,  $60\%$ ,  $70\%$  の場合の確率分布をプロットした。平均相関はそれぞれ  $\rho_{avg.} = 1.6\%$ ,  $3.7\%$ ,  $8.6\%$  となっている。期待値  $p$  は  $0.1\%$  にセットした。まず、確率密度関数はケーリーツリー上のモデルと同じように、ダウン数が小さい部分と、大規模な連鎖の部分に分離し、前者では急激に確率が減少しているが、後者では非常にゆっくりと減少する。相関が大きくなればなるほど、大規模な連鎖の確率の減少はゆるやかになり、指数分布的な振舞いをする事が分かる。そ

の分, ダウン数の少ない部分の確率は減少する. また, ケーリーツリーのモデルの確率分布で見られたような, 振動する振舞いは, このモデルの場合には確認できない. これらの振舞いは, サーバネットの連鎖ダウンを記述する二重指数分布と共通するものである.

#### 4. BA モデル上の相関のある二項分布, 二重指数分布

前章の結果から, スケールフリー性を持つ一般化ランダムグラフの場合, 相関のある二項分布は二重指数分布的な振舞いをすることが分かった. しかし, 一般化ランダムグラフはループのないネットワークなので, 現実のネットワークモデルとして, そこでの結論がどこまで適用度があるのかが不明である. そこで, 本セクションでは, BA モデルでの相関のある二項分布を解析する. 二重指数分布との比較も行う.

##### 4.1 BA モデル上の相関のある二項分布

確率モデルの構成方法から述べる. 基本的に一般化ランダムグラフの場合と同じように, 最初にネットワークの生成を行う<sup>9)</sup>. 一般化ランダムグラフの場合は, 同時分布を構成する際のノードの順序に同時分布が依存しないので, ノードの番号順に逐次条件付き確率を掛けて構成すればよかった. しかし, BA モデルの場合は, ネットワークがループを持つため, 同時分布がそれを構成するときのノードの順序に依存してしまう. そのため, さまざまな順序で同時分布を構成し, 順序に関する平均化を行う必要がある. 平均化の際の各順序の重みについては, すべての順序が同じ重みで寄与すると仮定する. 別の問題として, ネットワークがループを持つため, 条件付き確率のパラメータ  $\{\alpha_k, \beta_{k,i}\}$  を各変数ごとに決定するのが困難なことがある. 2次元正方格子のような単純かつ規則性のあるネットワークの場合なら, 適当にノードに順序を決めた場合,  $p, \rho$  という期待値, 相関を実現するパラメータを決定することはできる. 詳しくは付録の議論を参照されたい. しかし, BA モデルの場合は, スケールフリー性, ランダム性, ノードの順序がこうした計算を困難にする. そこで,  $z$  個の変数に接続するとき用いる条件付き確率のパラメータについて, 期待値の条件 ( $\langle X_i \rangle = p$ ) のみ課すことにする. 次のパラメータを用いる.

$$\alpha_z = \frac{p(1-\rho)}{(1-\rho) + z \cdot \rho}, \quad \beta_{z,i} = \frac{\rho}{(1-\rho) + z \cdot \rho} \quad (29)$$

これらのパラメータを用いた場合, 各変数の期待値は  $p$  となるが, 隣接する変数間の相関は必ずしも  $\rho$  とはならない. ただし, 接続相手の  $z$  個の変数が完全グラフの構造を持ち, かつ, すべてのペア間の相関が  $\rho$  となっている場合は, 新たに接続される変数と,  $z$  個の変

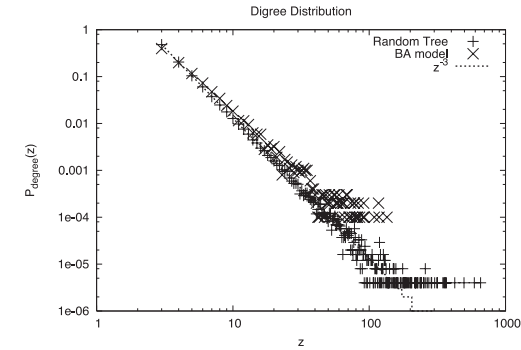


図7 一般化ランダムグラフ(+), BA モデル(x)の度数分布をプロットした. パラメータは  $N = 1000$  とした. 式(28)の分布関数も参考のためにプロット(点線)した

Fig. 7 Plot of the degree distributions of the Random Tree Model (+ symbol) and BA model (x symbol). We also plot the power law distribution Eq. (28). with dotted line.

数間の相関は  $\rho^{*1}$  となる.

以上の手順により, BA モデル上の相関二項分布を生成した. ただし, ツリー状のネットワークの場合と異なり, ループの存在のため確率分布関数の間に再帰的な関係式が存在しない. そのため, 条件付き確率に従って, モンテカルロ法で状態を生成し, それを用いて確率分布を求めた. サンプル数は  $10^5$  である.  $N = 10^3$  とし,  $10^2$  通りの順序と  $10^2$  通りのネットワークサンプルでの平均化も行っている.

まず, 確認のため度数分布をプロットしたのが図7である. 同じ図に, 前章で用いた一般化ランダムグラフの場合の度数分布と理論式(28)もプロットした. 一般化ランダムグラフの場合, 度数1の端点は分布から削除してある. BA モデルの度数は,  $\gamma = 3$  の逆べきの分布に従っていることが分かる.

次の図8は, BA モデル上の相関のある二項分布モデルの確率密度関数をプロットしたものである. 前セクションの一般化ランダムグラフの場合の確率密度関数もプロットしてある. 平均相関  $\rho_{avg}$  と隣接ノード上の変数間の相関  $\rho$  の関係を一般化ランダムグラフの場合と比較すると,  $\rho$  が同じなら BA モデルの方の平均相関のほうが, 一般化ランダムグラフのものより大きくなることが分かる. これは, BA モデルにはループがたくさん存在し, それらが相関をお互いに強化する方向に働くためである. モデルの構成方法のところでも解

\*1 このパラメータの選択はベータ二項分布を生成する確率過程に由来する<sup>5)</sup>.

### 33 ネットワーク上の相関のある二項分布

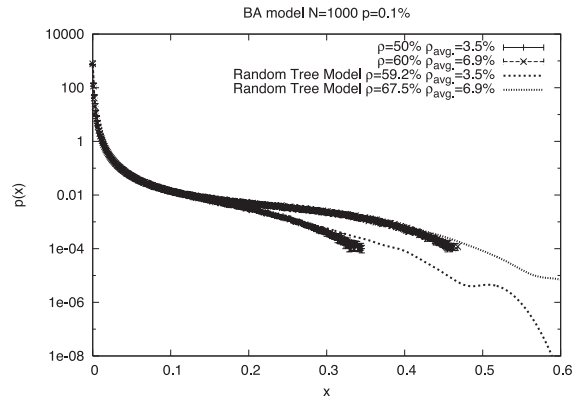


図 8 BA モデル上の相関のある二項分布の確率密度関数をプロットした。共通のパラメータは  $p = 0.1\%$ ,  $N = 10^3$  とし,  $10^2$  個のネットワークサンプルで平均したものである。隣接ノード間の相関には,  $\rho = 50\%$ ,  $60\%$  を用いている。一般化ランダムグラフの確率密度関数もプロットした

Fig. 8 Plot of  $p(x)$  for the BA model. We set  $p = 0.1\%$ ,  $N = 10^3$  in common. We take average over  $10^2$  network samples. About  $\rho$ , we use  $\rho = 50\%$ ,  $60\%$ . We also plot  $p(x)$  for the random tree model.

説したが, もしネットワークが完全グラフなら, その部分での  $\rho_{avg.}$  と  $\rho$  は等しくなり, 一方, ループのないツリー構造のグラフでは, ノード間の距離とともに相関は減衰するため,  $\rho_{avg.}$  は一般に  $\rho$  よりも小さくなる。BA モデルは完全グラフほどリンクが密集していないが, かといって一般化ランダムグラフのようなスカスカでもない。  $\rho$  が等しい場合, BA モデルの  $\rho_{avg.}$  は一般化ランダムグラフより大きくなるのである。

確率密度関数の振舞いは, 一般化ランダムグラフのものと同様に似た振舞いをしていことが分かる。大体  $\rho_{avg.}$  が  $3.5\%$  で揃っている, BA モデルで  $\rho = 50\%$  の場合と一般化ランダムグラフで  $\rho = 59.2\%$  の確率密度関数を比較してほしい。バルク部分では誤差の範囲内で, 確率密度関数は一致していることが分かる。同じく,  $\rho_{avg.}$  が  $6.9\%$  で揃っている, BA モデルで  $\rho = 60\%$  の場合と一般化ランダムグラフで  $\rho = 67.5\%$  の確率密度関数もバルク部分では一致している<sup>\*1</sup>。一方, 前者では  $x$  が  $0.3$ , 後者では  $0.4$  を超える端の部分では, 確率密度関数が一致していない。BA モデル上と一般化ランダムグラフ上の相関のある二項分

\*1 確率密度関数をモンテカルロ法で求める場合, 確率が小さい非常に稀な減少は, その出現回数のバラツキが平均出現数よりも大きくなってしまふ。グラフでは, 誤差が確率密度の絶対値を超えた部分のデータはプロットしていない。

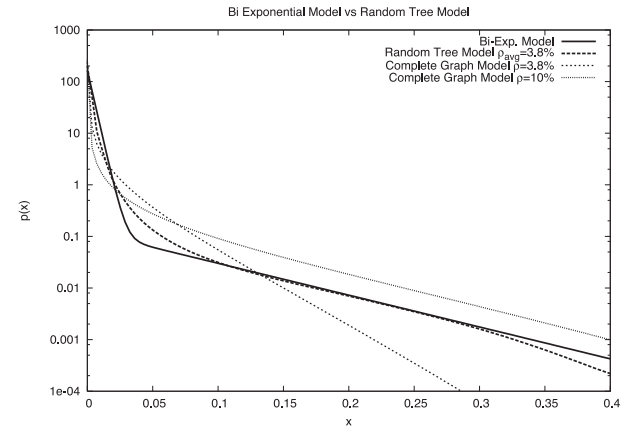


図 9 二重指数分布 (実線) と一般化ランダムグラフ上の相関のある二項分布の確率密度関数のプロット (点線)。前者のパラメータは  $\alpha = 0.009$  および  $r_1 = 0.4, r_2 = 0.95$ , 後者は  $p = 0.23\%$ ,  $\rho = 50\%$ ,  $\rho_{avg.} = 3.8\%$  とした。参考のため,  $\rho = 3.8\%$  と  $\rho = 10\%$  のベータ二項分布 (完全グラフ上の相関のある二項分布モデル) もプロットしている

Fig. 9 Plot of  $p(x)$  for the bi-exponential model (solid line) and correlated binomial model on the general random graph (dotted line). We set  $\alpha = 0.009$  and  $r_1 = 0.4, r_2 = 0.95$  and  $p = 0.23\%$ ,  $\rho = 50\%$ ,  $\rho_{avg.} = 3.8\%$ . We also plot the probability density for the beta binomial model with  $\rho = 3.8\%$  and  $\rho = 10\%$ .

布モデルの確率密度関数は完全には一致しないようである。理由としては, 一般化ランダムグラフの場合, 端のノードが多数存在してしまうので, 次数分布が 2 つのモデルで完全に一致していない。また, BA モデルでは, 隣接した確率変数間の相関が必ずしも  $\rho$ とはなっていない, など, いくつか理由をあげることができる。しかし, 両者の確率分布の類似性は高く, BA モデル上の相関のある二項分布モデルも二重指数分布モデル的な振舞いをするということが分かった。

#### 4.2 二重指数分布との比較

二重指数分布関数と一般化ランダムグラフ上の相関のある二項分布モデルとの比較を行う。ここで, より現実的な BA モデルではなく一般化ランダムグラフ上のモデルを用いた理由は, 両者の確率分布がデータと比較する部分  $0 \leq x \leq 0.2$  ではほぼ一致したからである。文献 1) の結果を用いると, PlanetLab という実験ネットワークでのサーバの連鎖ダウンの確率分布のパラメータは  $\alpha = 0.009$  および  $r_1 = 0.4, r_2 = 0.95$  となっている。サーバの平均台数は  $N = 277$ 。確率密度関数をプロットしたものが図 9 の実線である。同じ図に, 一般

化ランダムグラフ上の相関のある二項分布の確率密度関数もプロットした。  $p = 0.23\%$  で、  $\rho = 50\%$  ,  $\rho_{avg.} = 3.8\%$  の場合である。

ここで、グラフの表示範囲は  $0 \leq x \leq 0.4$  としている。理由は、文献 1) のサーバの連鎖ダウンの確率分布のフィットにおいて、データが  $x \leq 0.2$  に限られているからである。ここでは、平均 277 台のサーバ中、観測された最大の連鎖ダウンのサイズは 60 台弱であった。図の確率分布を見比べると、ほぼ同じような振舞いをしていることが分かる。細かい点をいえば、  $x = 0.03$  での関数の曲がり方が、一般化ランダムグラフ上のモデルのほうが滑らかであるとか、  $x \geq 0.3$  では、2 つのカーブが一致しなくなるなど、相異点をあげられる。前者に関しては、実データもカーブの部分では二重指数分布モデルとは一致せず、むしろ一般化ランダムグラフのモデルの振舞いに近いように見える。後者については、最初に述べたように、そのサイズの連鎖ダウンは確認されていないので、一致しなくても問題はない。しかし、  $\gamma = 3.0$  のベキの次数分布を持つ、期待値が  $p$ 、隣接変数間の相関が  $\rho$  という非常に単純なモデルでもここまで実際のデータと一致するという事は興味深い。一方、同じグラフにベータ二項分布の確率密度関数もプロットした。ベータ二項分布は、  $\rho$  を大きくすると、大規模連鎖を記述する部分の傾きが小さくなり、二重指数分布と似た振舞いを示す。しかし、ダウン数が少ない部分の確率が減少しすぎで、二重指数分布と重なるようにパラメータを決めることはできない。  $\rho = 10\%$  の場合に、  $p$  をさらに小さくして大規模連鎖の部分に二重指数分布にフィットさせると、ダウン数が少ない部分の確率はさらに小さくなり、データとの乖離が激しくなる。

## 5. まとめと今後の課題

本論文では、ネットワーク上の相関のある二項分布の 1 つのモデルを提案した。ノード上の変数の期待値、隣接するノード上の変数間の相関係数を与え、ノードを接続するごとに条件付き確率を用いて同時分布関数を構成するものである。得られた同時分布関数がノードの接続の順序に依存しないための十分条件として、(1) ネットワークがツリー状のループを持たないこと、(2) ループが存在する場合でも、それを含む最小部分が完全グラフを構成し、完全グラフ上のすべての変数の期待値が等しく、かつすべての変数間の相関係数が等しいこと、以上の 2 つをあげた。一般のネットワークの場合は、順序を決め、かつ単純な構造の場合、与えられた期待値、相関を実現しながら同時分布を構成することは可能であるが、複雑なネットワークの場合は、それを実行するのが困難であることも述べた。この構成方法をもとに、1 次元格子、ケーリーツリー、一般化ランダムグラフ、BA モデル上の相関のあ

る二項分布を構成し、その確率分布関数の振舞いを調べた。特に、後者のスケールフリー性を持つ 2 つのモデルでは、二重指数分布と似た振舞いをすることも示した。このことから、サーバネットワークのダウン数の確率分布が二重指数分布モデルで記述できる理由の 1 つが、ネットワークのスケールフリー性（次数分布がべき則に従うこと）であること、また、データはダウン相関が非常に強いことを示唆することを述べた。

今後の課題としては、スケールフリー性を持つモデルでの確率分布関数がなぜ二重指数分布と似た形になったのかを解析的に明らかにすることがある。1 次元の場合は、期待値が小さく相関が大きい極限で指数分布が出てくることを示したが、それ以外のモデルでは数値的に求めただけで、理解が十分深まったとはいえない。確率分布の形とネットワークの形の間関係の理解をさらに深める必要がある。一般化ランダムグラフ上のモデルと BA モデル上の相関のある二項分布モデルの確率分布がともに二重指数分布と似た振舞いを示したことは、確率分布はネットワークの次数分布で大体決まるように思われる。この点を解析的に明らかにしたい。また、今回、二重指数分布というモデルを目標に、ネットワーク上の相関のある二項分布モデルの研究を進めてきたが、もともと二重指数分布はサーバの連鎖ダウンのモデルであった。では、サーバの連鎖ダウンは、本論文で述べたようなモデルで記述できるのかも重要な問題である。たとえば、相関関数の指数関数的な振舞いを現実のデータは示すのか、ネットワーク構造やサーバのダウン率などを正確にインプットした場合、連鎖ダウンのデータとより精度良く一致するのか、といった問題がある。そのためには、WEB サーバ、ストレージサーバ、携帯電話の基地局のネットワークなど、相互依存していると思われる系のデータを集め、解析する必要がある。こうした研究は、より安全なネットワークの設計の際、重要な指針を与えるはずである。特に、Nath らの論文では<sup>1)</sup>、ネットワークの設計指針を導く際に、ダウン数は二重指数分布モデルに従うが、ダウンするサーバをランダムに選ぶと仮定したモデルを用いている。これは、ネットワークが完全グラフと仮定し、ネットワーク構造を完全に無視することに対応する。我々の提案する方法は、どのようなネットワーク構造を採用した場合にどのような確率分布が得られるのかを具体的に計算できるので、リスク評価をより精度良く行えると考えているが、今後の課題である。また、連鎖の確率のネットワーク構造依存性の解明は、連鎖デフォルトを記述する確率モデル<sup>13)</sup>を構成する際にも重要になると考えている。

謝辞 One of the authors (S.M.) thanks Dr. E. Guitter and Dr. Ph. Di Francesco for useful discussions and suggestions.

## 参 考 文 献

- 1) Nath, S., Yu, H., Gibbons, P.B. and Seshan, S.: Subtleties in Tolerating Correlated Failures in Wide-area Storage Systems, *Proc. 3rd conference on Networked Systems Design & Implementation*, pp.225–238 (2006). downloadable at <http://www.usenix.org/events/nsdi06/tech/nath.html>
- 2) Griffiths, D.A.: Maximum Likelihood Estimation for the Beta-Binomial Distribution and an Approach to the Household Distribution of the Total Number of Cases of Disease, *Biometrics*, Vol.29, No.4, pp.637–648 (1973).
- 3) Williams D.A.: The Analysis of Binary Responses from Toxicological Experiments involving Reproduction and Teratogenicity, *Biometrics*, Vol.31, No.4, pp.949–952 (1975).
- 4) Kupper, L.L. and Haseman, J.K.: The Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments, *Biometrics*, Vol.34, No.1, pp.69–76 (1978).
- 5) Hisakado, M., Kitsukawa, K. and Mori, S.: Correlated Binomial Models and Correlation Structures, *J. Phys. A.*, Vol.39, No.50, pp.15365–15378 (2006).
- 6) Schönbucher, P.J.: *Credit Derivatives Pricing Models*, New York, Wiley (2001).
- 7) Hull, J. and White, A.: Valuing Credit Derivatives Using an Implied Copula Approach, Working Paper, University of Toronto (2006). downloadable at <http://www.defaultrisk.com>
- 8) Newman, M.E.J., Strogatz, S.H. and Watts, D.J.: Random Graphs with arbitrary degree distributions and their applications, *Phys. Rev. E*, Vol.64, pp.26118–26134 (2001).
- 9) Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol.286, pp.509–512 (1999).
- 10) 宮川雅巳: グラフィカルモデリング, 朝倉書店 (1997).
- 11) Konno, N.: Limit Theorems and Absorption Problems for One-Dimensional Correlated Random Walks, preprint arXiv:quant-ph/0310191.
- 12) 今野紀雄: 量子ウォークの数理, 産業図書 (2008).
- 13) Sakata, A., Hisakado, M. and Mori, S.: Infectious Default Model with Recovery and Continuous Limit, *J. Phy. Soc. Jpn.*, Vol.76, No.5, pp.54801–54807 (2007).

## 付録 二次元正方格子上の相関のある二項分布の構成

本文では, ループを持つネットワークでも規則的な簡単な格子なら期待値  $p$ , 隣接相関  $\rho$  の条件を満たす相関のある二項分布を構成できると述べた. その具体的な例として, 二次元正方格子上のモデルの構成方法を説明する.

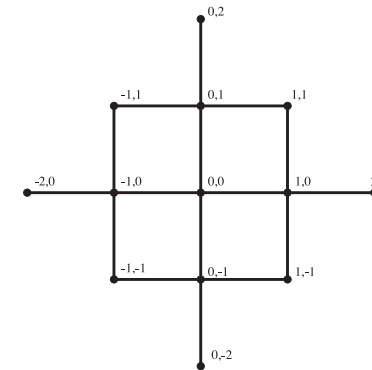


図 10  $N = 13, L = 2$  の 2 次元正方格子  
Fig. 10 Picture of two dimensional lattice.  $N = 13, L = 2$ .

まず, 二次元正方格子のノード (格子点) の順序を決める必要があるので, 二次元正方格子を作成する. まず, 原点にノードをおき, それを  $(0, 0)$  でラベルする. 次に, 原点から距離 1 の 4 個のノードをその座標  $(1, 0), (0, 1), (-1, 0), (0, -1)$  で指定する. 順序は,  $(1, 0)$  から反時計回りの順番とする. 同様に, 原点から距離 2 の 8 個のノードをその座標  $(i, j)$ ,  $|i| + |j| = 2$  で指定し, 順序は  $(2, 0)$  から反時計回りの順番. このプロセスを繰り返して, 原点から距離  $L$  までの正方格子を用意し, そのノードに順序を決める. ノードの総数  $N_L$  は

$$N_L = 1 + 2L(L + 1)$$

である.

次に, 同時分布関数を構成する. まず, 原点の変数の確率分布は

$$p(x_{0,0}) = p^{x_{0,0}} \cdot q^{1-x_{0,0}}$$

である. これに, ノード  $(1, 0)$  を接続するとき, 式 (20) を用いて

$$p(x_{0,0}, x_{1,0}) = p(x_{1,0}|x_{0,0}) \cdot p(x_{0,0})$$

とする. ノード  $(0, 1), (-1, 0), (0, -1)$  も同様.

次に距離 2 のノードに関しては, 1 個のノードにしか依存しない  $(2, 0), (0, 2), (-2, 0), (0, -2)$  の場合は式 (20) を用いて接続する. 2 個のノード上の変数  $y, z$  に依存する場合は次の条件付き確率を用いる.

### 36 ネットワーク上の相関のある二項分布

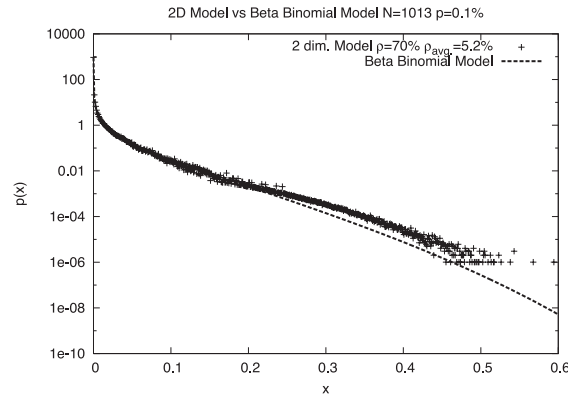


図 11 二次元正方形格子とベータ二項分布の確率密度関数のプロット

Fig. 11 Plot of  $p(x)$  for the two dimensional lattice model and the beta-binomial model.

$$p(x|y, z) = (\alpha_2 + \beta_2(y + z))^x \cdot (\alpha_2' + \beta_2((1 - y) + (1 - z)))^{1-x}$$

$$\alpha_2 = p \frac{(1 - \rho)^2}{1 + \rho^2}, \quad \alpha_2' = q \frac{(1 - \rho)^2}{1 + \rho^2}, \quad \beta_2 = \frac{\rho}{1 + \rho^2} \quad (30)$$

たとえば、ノード  $(1, 1)$  は、 $(1, 0)$  と  $0, 1$  に依存する。このとき用いる条件付き確率は、 $p(x_{1,1}|x_{1,0}, x_{0,1})$  である。

この系の相関関数は、たとえば第 1 象限で次の漸化式に従うことが分かる。

$$\rho_{(i,j),(k+1,l)} = \rho_{(i,j),(k,l+1)} = \rho \cdot \rho_{(i,j),(k,l)}$$

$$\rho_{(i,j),(k+1,l+1)} = \beta_2 \cdot (\rho_{(i,j),(k,l+1)} + \rho_{(i,j),(k+1,l)})$$

たとえば、原点  $(0, 0)$  とノード  $(1, 1)$  の間の相関は

$$\rho_{(0,0),(1,1)} = \beta_2 \cdot (\rho + \rho) = \frac{2\rho^2}{1 + \rho^2}$$

である。この漸化式から、ノード  $(i, j)$  と  $(k, l)$  の間の相関は、距離  $|k - i| + |l - j|$  とともに指数関数的に減衰することが分かる。

このモデルの確率分布関数をプロットしたのが図 11 である。ノード数  $N$  は  $N = 1013$ 、原

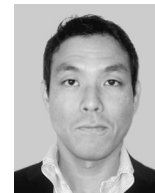
点から端点までの距離  $L$  は  $L = 22$  とした。モンテカルロ法でサンプリングし、確率密度関数を求めている。サンプル数は  $10^6$ 。ほぼベータ二項分布と同じように振る舞うが、裾野が少し広がり厚みを増している。

ここで構成した確率分布について注意を述べる。本文でも解説したように、ノードの順序のとりかたによって、同時分布は一般に異なる。ここでは、1 つの順序に対して同時分布を構成し、その性質について述べたが、本来はあらゆる順序について平均化する必要がある。そのため、ここでの結果がそのまま正方形格子上の相関二項分布について成立する保証はない。しかし、十分大きな系を考えた場合、そのマクロな性質が微細な順序という情報にあまり依存しない可能性もある。今後の研究課題である。

(平成 20 年 1 月 11 日受付)

(平成 20 年 6 月 16 日再受付)

(平成 20 年 7 月 31 日採録)



守 真太郎

昭和 42 年生。平成 8 年東京大学大学院理学系研究科物理学専攻博士課程修了。平成 7 年から平成 8 年学術振興会特別研究員。平成 8 年フランス SACLAY (SPhT) 客員研究員。平成 9 年より北里大学理学部講師。高分子物理学、スピン系の統計物理学、反応拡散系の場の理論に関する研究に従事。理学博士。日本物理学会会員。



久門 正人

昭和 43 年生。平成 8 年東京大学大学院理学系研究科物理学専攻博士課程修了。平成 9 年から 10 年学術振興会特別研究員。平成 11 年から 12 年、金融庁勤務。平成 13 年よりスタンダードプアーズのクオンツ・アナリスト。可積分系、数値物理学、金融工学に関する研究に従事。理学博士。