

企業の倒産とインターネットサーバーのダウンの確率分布と相関の推定

Statistical Properties of Correlated Failures and Defaults

守 真太郎

Mori Shintaro

北里大学理学部物理学科

Department of Physics, School of Science, Kitasato University

mori@sci.kitasato-u.ac.jp, <http://sharaku.sci.kitasato-u.ac.jp/mori/>

久門 正人

Hisakado Masato

スタンダード&プアーズ

Standard and Poors

masato_hisakado@standardandpoors.com

キーワード: default, implied, correlation, network, failure

概要

企業の連鎖倒産とインターネット上のサーバーの連鎖的なダウンの確率分布関数の性質について報告する。前者については、iTraxx-CJという連鎖倒産にたいするクレジットデリバティブ(保険)のプレミアムをもとに確率分布を推定。後者については、S.Nath et al らの3つのサーバーネットワークの同時ダウンの解析結果をもとにする。0,1タイプの確率変数の従う相関のある確率分布についての我々の結果(完全グラフ、1次元およびケーリー木などの空間上の確率分布の構成方法、解析方法)を用い、企業ネットワークの連鎖倒産やサーバーネットワークの同時ダウンの連鎖の強さを推定する。

1. はじめに

企業の連鎖倒産やサーバーネットワークの連鎖ダウンなどのシステミックリスクは、現在まで盛んに研究されてきた。たとえば、インターネットは次数が逆べきの分布に従うというスケールフリー性を持ち、ランダムなノードのダウンに対しては、ネットワークの接続性は頑健だが、次数のおおきなハブの攻撃には脆弱である [1, 2, 3]。企業の倒産についても、債務不履行(デフォルト)のリスクを正確に評価し、ヘッジするための手法の研究が行われている [4]。倒産の数の時間変化を見ると、多い年が連続したりというクラスタリングが見られ、企業は独立に倒産するのではなく、依存性(相関)がある。また、サーバーネットワークやルーターのネットワークについても大規模なルータ網のダウン、IP電話ネットのダウンなど、連鎖的なダウンが頻発している。つまり、サーバーやルーターのダウンには相関があり、独立に起こる場合だけではない。しかし、実際のサーバーのダウンのデータから、その相関や依存性の詳細な研究が始まったのは、ここ数年のことである。たとえば、Bakkalogluらは、10台程度のコンピュータネットワークの連鎖ダウンの研究から、連鎖の強さを条件付確率で表すモデルを提案した [5]。また、PLANET Lab やウェブサーバーのダウンの詳細なデータから、ダウンの数の確率分布が非常にファットテールなテールをもつことが指摘された [6, 7]。ファットテール

ルになる原因として、ウイルスの蔓延やDDoS攻撃などのアタック、人為的なミスにより引き起こされたからである。では、そこにはどのような確率法則があり、また連鎖の強さはどれぐらいだと評価できるのか?

こうした確率法則の解明、連鎖の強さの評価は、HOT[3]などで、システムを設計する場合の不可欠の情報である。一方、企業の連鎖倒産に関しては、クレジットデリバティブ市場におけるCDO(Collateralized debt Obligation)の取りひきの開始が、連鎖倒産、デフォルト相関の研究の引き金となった。日本においてはCDOの一種であるiTraxx-CJの取引が始まり、そのプレミアムを説明するための確率模型の研究は現在でも盛んに行われている [8]。では、どのような確率模型がデフォルト連鎖を記述し、どれぐらいの連鎖の強さがあるとマーケット参加者は見込んでいるのか?

今回の発表では、クレジットデリバティブの世界で想定されている連鎖倒産やサーバーネットワークの同時ダウンの分布関数の背後にある確率法則に焦点を当て、ネットワーク上の確率分布の構成方法、解析方法をもとに連鎖の強さの推定を行う。

2. デフォルト連鎖の強さを推定する

iTraxx-CJ について述べる前に、クレジットデリバティブ市場においてもっとも重要な商品であるCDS(Credit

Default Swap) について説明する。CDS は企業の債務不履行 (デフォルト) に対する保険商品であり、デフォルトリスクに見合った金利 (プレミアム) を払うと、デフォルトの際、損失をカバーしてくれるというものである。そのプレミアム s は、対象とする企業の債権のデフォルト確率 P_d とデフォルトが起こった場合、額面のうちのどれだけが戻るのかという回収率 R および、満期期間 T により、次の近似式で結ばれている。

$$P_d \simeq \frac{s \cdot T}{1 - R}$$

満期 T は 5 年で、回収率は 35% である。たとえば、2005 年の 7 月 5 日において、JAL の CDS のプレミアムは 88.111[BPS]、TOYOTA は 8.182[BPS] で (BPS とは、100 分の 1%)、デフォルト確率はそれぞれ 6.4%、0.61% と評価することが出来る (この結果は近似式ではなく、無リスク金利などを考慮したものである)。ポイントは CDS のプレミアムから市場の予測する企業のデフォルト確率 P_d が推定できることである。つまり倒産確率についてマーケットの考えがプレミアムから読めるわけである。しかし、この情報だけでは連鎖倒産についてクレジット市場の考えを読むことはできない。

そのためには、こうした CDS を多数束ねたポートフォリオを考える必要がある。そうした商品は実際に存在し、一般に CDO と呼ばれ、デフォルト可能性のある債券 (CDS など) のポートフォリオに優先劣後構造を入れたものである。(CDS をベースとした CDO のことをシンセティック CDO と呼んでいる。) 日本では、iTraxx-CJ がマーケットで日々取引が行われ、さまざまな CDO の中で基準となっている。トヨタ、JAL、ソニー、武富士、鹿島など、日本の企業 50 社の CDS のポートフォリオをもとにして構成されている。そのプレミアムは対象とするリスク部分に応じて 6 つある。まず、ポートフォリオ全体のリスクに対する保険はインデックスといい、そのプレミアムからポートフォリオ全体の平均の損失率、デフォルト確率を推定することができる。2005 年 7 月 5 日のインデックスの値は 24.553[BPS] で、この値から平均のデフォルト確率は $P_d = 1.83\%$ 。2005 年 8 月 31 日だと、インデックスは 22.08[BPS] で、 $P_d = 1.64\%$ と読むことができる。つまり、5 年後までに倒産する会社は平均 1 社を下回る程度。それが、クレジットマーケットの推定する 5 年後の状況ということになる。しかし、インデックスだけでは、5 年後の状況について、平均の倒産数しかわからないため、連鎖倒産がどの程度起こると考えているのか、つまり倒産の間の相関についての情報を得ることができない。もし、倒産に相関がないなら、5 年後の倒産件数 n の従う確率分布 $P_{50}(n)$ は、項数 $N = 50$ 、確率 $p = P_d$ の 2 項分布 $\text{Bi}(50, P_d)$ となるはずだが、本当にそうなのだろうか？

確率分布 $P_{50}(n)$ について、 n の平均値以上の情報を教えてくれるのが、ポートフォリオを優先劣後構造により

分割 (トランシングと呼ぶ) したトランシェのプレミアムである。デフォルトが起こったときに、劣後トランシェからその損失を吸収し、そのトランシェが消滅したらその次により優先度の低いトランシェが損失をカバーするという仕組みになっている。各トランシェは、アタッチメント点 AP とデタッチメント点 DP をもち、発行時の額面は

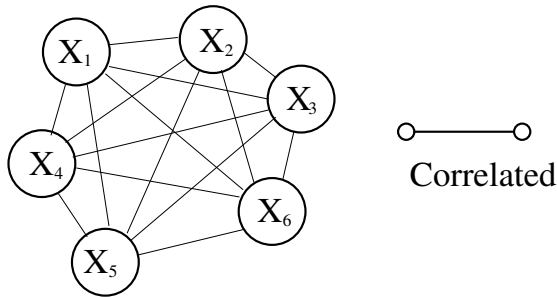
$$N_0 = 50 \cdot (\text{DP} - \text{AP})$$

である。損失が AP に達したらそのトランシェが損失を吸収し始め、DP をこえるとトランシェはなくなる。iTraxx-CJ は、5 つのトランシェを持ち、優先順位の低いものから順に {AP%, DP%} (名前) を書くと、{0%, 3%} (エクイティー)、{3%, 6%} (メザニン 1)、{6%, 9%} (メザニン 2)、{9%, 12%} (シニア 1)、{12%, 22%} (シニア 2) となっている。そして、各トランシェのプレミアムは、各トランシェの額面が満期までにどの程度毀損し、減少すると期待されるのかに依存してきまる。1 件のデフォルトで $1 - R = 0.65$ だけポートフォリオは価値を失う。すると、エクイティーは 3% の厚さを持っているので、デフォルト前の額面 N_0^1 は $N_0^1 = 1.5$ で、最初の 2 件のデフォルトでは額面は残るが、3 件デフォルトすると消滅し、エクイティーでカバーしきれなかったロスの残り 0.45 は、次のメザニン 1 が引き受けることになる。そして、4 件目までのデフォルトはメザニン 1 がカバーするが、5 件目からはメザニン 2 にも損失が生じる。すると、メザニン 1 のプレミアムは、3 件以上のデフォルトの起こる確率が決定し、メザニン 2 は、5 件以上のデフォルトの確率が関与する。一方エクイティーは 3 件以上の確率も必要だが、そのプレミアムの決定には 3 件未満のデフォルトの確率も必要であることがわかる。このように、各トランシェは、その AP、DP に対応して、どのリスクをカバーするのが決まり、またそのリスクの確率から額面がどれくらい変化するか期待値がわかり、プレミアムが決まることになる。逆に言えば、各トランシェのプレミアムがわかれば、そのトランシェが満期においてどの程度になるのかがわかり、担うリスクの確率を評価することが出来ることになる。次の表は、2005 年 8 月 30 日のインデックス および各トランシェのプレミアムと、各トランシェの担う額面 (全体で 50) N_0^i 、満期における期待額面 $N_{T, Implied}^i$ を示したものである。エクイティーに関しては、通常のプレミアムのほかに、最初に保険契約者が保険を売る側に払うアップフロントも () の中に示した。

では、以上の 5 つのプレミアムや期待額面からデフォルト数の確率分布を推定することは出来ないだろうか？ もちろん、確率分布は 50 個の CDS のデフォルトに関するものであり、さまざまな対称性を仮定しても自由度は 5 よりずっと大きい。つまり、5 つのプレミアムとインデックスの情報から確率分布を一意に決める方法はな

表 1 iTraxx-CJ の 2005 年 8 月 30 日の各トランシェのプレミアムと初期の額面、プレミアムから求めた満期の期待額面。

Tranche (i)	s_i [bps]	N_0^i	$N_{T,Implied}^i$
エクイティ (1)	300(1313.3)	1.5	1.1066
メザニン 1 (2)	89.167	1.5	1.4361
メザニン 2 (3)	28.5	1.5	1.4792
シニア 1 (4)	20.0	1.5	1.4854
シニア 2 (5)	14.0	5.0	4.9660
インデックス (6)	22.08	50	49.464



Exchangeable and not independent .

図 1 企業のデフォルト間の依存構造をグラフで表したものの。完全グラフの依存関係を仮定してデフォルト分布関数を推定する。グラフの頂点が CDS の状態を表し、リンクは相関のあることを意味している。

い。そこで、Hull と White は、CDS が交換可能である (つまり区別できない) と仮定し、デフォルト数分布を 2 項分布の重ねあわせで表現し、重ね合わせの関数がなめらかという条件を用いて確率分布を推定する方法を提案した [8]。つまり、各 CDS の状態を 0, 1 型の確率変数 X_i ($i = 1, 2, \dots, 50$) であらわした場合、次の図 1 の完全グラフで依存関係を表現することになる。ここで、 $X_i = 1$ は、デフォルト状態、 $X_i = 0$ はノンデフォルト状態を表す。

一方、我々は同じく交換可能であるとして自由度を 50 にまで落とし、さらにエントロピー最大原理によりデフォルト数の確率分布を推定する方法を提案した [9]。ここで、混合関数のなめらかさの代わりにエントロピー最大原理を用いた理由は、インデックスの情報だけでエントロピーを最大にすると 2 項分布となり、さらにトランシェの情報を加えていくと 2 項分布がゆがんで、デフォルト相関の情報が入ると考えたほうが自然だと考えたからである。Hull らのように混合関数になめらかさを要請すると、インデックスの情報だけだと 2 項分布より幅の広い分布となり、デフォルト相関がなめらかさというパラメータに依存してしまう。もちろん、2 項分布になるということは、インデックスの情報だけだとデフォルトが独立に起こるとすることを暗黙のうちに仮定しているので、恣意性がまったくないわけではない。その意味で、Hull & White の方法と我々の方法で解析し、それぞれの結果を比較吟味する必要があると考えられる。しかし、

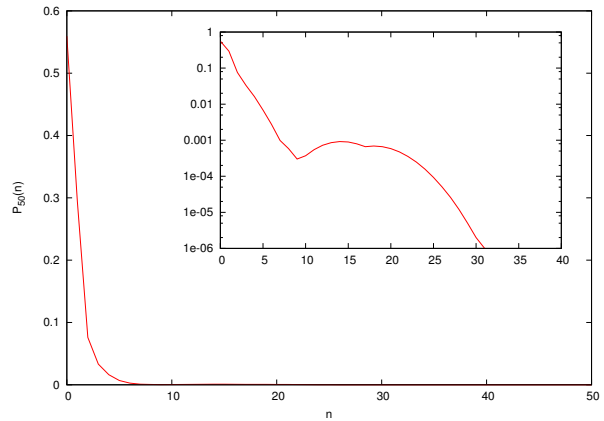


図 2 2005 年 8 月 30 日のプレミアムから推定した 5 年後の倒産数の確率分布関数 $P_{50}(n)$ をプロットしたものの。横軸はデフォルト数 n 。

交換可能性を課し、期待額面の条件を追加すると、確率分布を变形する自由度はかなり局所的な限定されたものになるので、両者の方法の差は少ないと予想している。

図 2 が 5 年後の企業のデフォルト数の確率分布である。このような、マーケットのプレミアムから推定したデフォルト確率分布のことをインプライドデフォルト分布と呼ぶ。グラフから分かるのは、デフォルト数が増加すると確率は急激に減少するが、あるところで確率の減少はとまる。そして 10 件から 20 数件のデフォルトに対して 0.1% から 0.01% のオーダーの確率が存在する。独立なデフォルトを仮定した 2 項分布に比べ、かなりの連鎖倒産の確率をマーケットはトランシェのプレミアムに含めていることが分かる。

では、倒産の連鎖の強さはどのぐらいなのだろうか？我々は、交換可能な 0, 1 の値をとる確率変数の同時分布について、その条件つき期待値 p_{ij} と条件付相関 ρ_{ij} のみならず必要条件を導いた [10]。ここで、添え字の ij は、 N 個の変数のうち i 個が 1、 j 個が 0 の条件を意味する。必要条件は、これらの条件付期待値、相関に対して次の式で与えられる。

$$p_{i+1j} = p_{ij} + (1 - p_{ij})\rho_{ij} \quad (1)$$

$$p_{ij+1} = p_{ij} - p_{ij}\rho_{ij} \quad (2)$$

$$p_{i-1j} - p_{ij-1} = -(1 - p_{i-1j})\rho_{i-1j} - p_{ij-1}\rho_{ij-1} \quad (3)$$

こうした条件を満たす条件付期待値、相関が与えられると、確率分布は一意に構成することができる。逆に、実際の確率分布が与えられると、条件付期待値、相関を求める手法も開発した [9]。

その方法を手短かに説明すると、図 3 のパスカルの三角形の斜辺に条件付確率 p_{ij} 、頂点に相関 ρ_{ij} を置き、底辺の同時分布関数 X_{N-n}^n から上向きに必要条件を解いていくことにより、 p_{ij} 、 ρ_{ij} を求めていく。同時確率分布

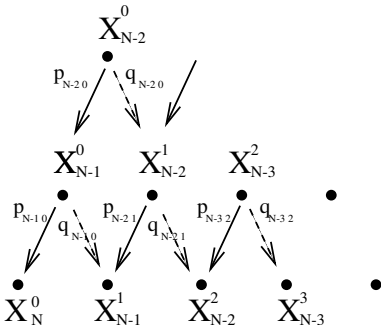


図 3 パスカルの三角形による相関のある 2 項分布の構成。

X_{N-n}^n は、確率分布 $P_N(n)$ から次の関係式で求める。

$$X_{N-n}^n = \frac{P_N(n)}{NC_n}$$

パスカルの三角形の頂点に達すると、無条件の期待値 p_{00} 、相関 ρ_{00} となり、これが平均デフォルト確率 P_d と相関 ρ_d である。もっとも、平均デフォルト確率、相関を求めるだけなら、インプライド分布から 1 次、2 次のモーメントを計算して求めることも出来る。求まったデフォルト相関（連鎖倒産の強さ）を次の表 2 に示した。

この表で、最初の 5 つの相関の値は、コピュラ模型でデフォルト確率分布を記述した場合に、プレミアムから逆算したときの相関の値である。トランシェ毎に相関の値がまったく異なっている。この現象はコリレーションスマイルと呼ばれ、コピュラがデフォルト確率分布をまったく再現できないからだとされてきた [8]。最後の相関の値がインプライド分布から求めたデフォルト相関の値である。コピュラによる相関とは異なり、分布をプレミアムから決め、相関を求めているので、相関の値が一意に定まる。そして、マーケットは 6.55% 程度の連鎖を推定していることが分かった。このように、デフォルト相関の値が一意に定まったことは重要で、この値から 1 件デフォルトした場合、次の CDS のデフォルト確率を $P_d + \rho_d(1 - P_d)$ の式で求めることが出来る。すると、最初の 1 件のデフォルト確率は 1.64% だが、次のデフォルト確率は約 8% と評価でき、関連するクレジットデリバティブの評価を行うことが可能になる。

図 4 には、同じデフォルト確率、相関 P_d, ρ_d を持つさまざまな確率模型のデフォルト分布関数とインプライド分布をプロットした。同じデフォルト確率 P_d とデフォルト相関 ρ_d を持っても、分布関数の振る舞いが全く異なる。では、これらの振る舞いの差はどのように理解すればよいのか？そこで、 ρ_{n0} と p_{n0} の振る舞いに着目してみる。

1	2	3	4	5	インプライド
13.5	1.20	2.58	4.95	9.71	6.55

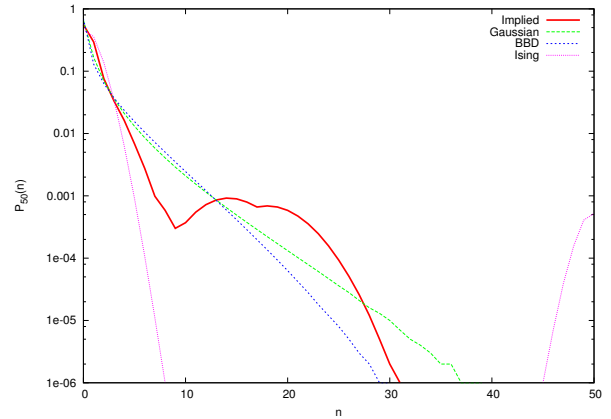


図 4 インプライド分布（赤）とさまざまな確率模型の分布関数をプロットした。デフォルト確率 P_d 、デフォルト相関は ρ_d はインプライド分布と同じ値にしてある。ガウス型コピュラが緑線、ベータ 2 項分布が青線、イジング模型が紫の線。

次の図 5 がデフォルト数 n の関数として ρ_{n0}, p_{n0} をプロットしたものである。まず、インプライド分布の ρ_{n0} の振る舞いを見てみると、それは平均のデフォルト相関 $\rho_{00} = \rho_d$ から始まり、デフォルト数 n とともに相関が増加し、20% 弱の相関を持つに至る。ピークののち、相関は急激に減少し、7 件程度でほぼゼロになる。つまり、マーケットは 7 件程度のデフォルトが起こると、無相関となり、それ以後はデフォルトは独立に起こると推定していると考えられる。一方、ベータ 2 項分布やコピュラ模型では、低いピークをもとことはあっても、ほぼ緩やかに減少し、ゆっくりとゼロに近づく。つまり、デフォルトがある程度おこっても、デフォルト相関が消えることはない。イジング模型の場合は、インプライド分布と同様に、 n とともに急激に増加し、90% 近くまで到達した後、急激にダンプする。こうした、 ρ_{n0} の振る舞いは、次の式 (1) と同じもの) で p_{n0} の振る舞いに反映する。

$$p_{n+10} = p_{n0} + (1 - p_{n0})\rho_{n0}$$

この漸化式から、図 5 の p_{n0} の振る舞いを理解することができる。 p_{n0} は、 $n = 0$ で、 $p_{00} = P_d$ から出発し、 $n = 1$ において $p_{10} = P_d + (1 - P_d)\rho_d$ を通過する。この点までは、同じ P_d と ρ_d をもつ模型なら同じ振る舞いである。問題は $n \geq 2$ における振る舞いである。イジング模型の場合は、 ρ_{n0} が急激に増加し、1 に近づくので、対応して p_{n0} もすぐ 1 になる。すると、確率分布は、 $p_{00} = P_d$ と $p_{n0} \approx 1$ のふたつの確率の 2 項分布を重ね合わせたものとなる。実際、イジング模型の確率分布関数 $P_N(n)$ は、原点近傍 $n \approx 0$ の大きなピークと、全滅を表す $n = 50$ の小さなピークのふたつの 2 項分布の重ね合わせとなっている。一方、ベータ 2 項分布や、コピュラ模型では、 ρ_{n0} がゆっくりゼロに近づくのに対応して、 p_{n0} もゆっくり 1 に近づき、結果としてさまざまな p の 2 項分布の重ねあわせることになる。確率分布も、裾野がゆっくりとダンプするなめらかな振る舞いを持っている。一方、イ

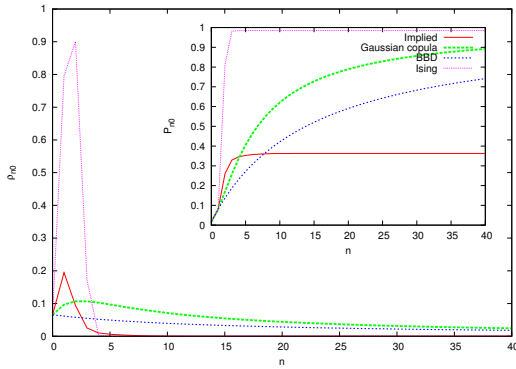


図 5 条件付相関 ρ_{n0} と期待値 p_{n0} (inset figure) をプロットした。確率分布、模型とグラフの対応は前の図と同じ。

インプライドは、 ρ_{n0} がイジングに比べれば低いピークを持った後、急激にダンプする。そのため、 p_{n0} が 1 になるまえに p_{n0} の増加はとまり、図の場合だと約 35% のところで一定となる。 $P_N(n)$ のほうも、 $N \cdot 0.35 \approx 18$ のところに、2 個目の 2 項分布のピークを持った形をしている。

まとめると、

- イジング模型は ρ_{n0} が急激に 1 に増加し、その後急激にダンプするため、確率分布は $p = P_d$ の大きなピークと全滅のところに小さなピークをもつ。イジング模型は「ほとんど独立な小規模な連鎖倒産とほぼ全滅の大規模な連鎖倒産」を記述する模型である。
- ベータ 2 項分布やコピュラ模型では、 p_{n0} がゆっくりと 1 に近づくので、確率分布はさまざまな大きさの p の 2 項分布が混ざり合ったものとなる。小規模な連鎖倒産から大規模な連鎖倒産まで、いろいろな規模の連鎖倒産を記述する模型である。
- インプライド分布では、 ρ_{n0} がイジングと似た振る舞いをするが、 p_{n0} が 1 に近づくまえに相関が消滅し、 p_{n0} が一定となる。そのため、マーケットは小規模なほとんど独立な連鎖倒産と中規模な連鎖倒産を予想していると考えられる。

では、このような特徴をもった確率模型は存在するのか？ベータ 2 項分布やコピュラ模型では、なめらかにさまざまな規模の連鎖倒産が起こるような模型であるため、インプライド分布とはかなり構造がことなり、CDO のプレミアムを説明することはできない。むしろ上で見たように、イジング模型と非常に似ている [11, 12]。最近の研究では Lo and Davis の感染模型 [13] を拡張するとプレミアムを非常によく再現できるようにパラメータが選べる [14] ことも分かってきたが、決定的な確率模型は存在しない。iTraxx-CJ などの CDO、シンセティック CDO の市場がおおきくなりつつある現在そうした確率模型を開発する必要がある。(もっとも、現在の主流はインプライド分布の情報をベースにして、他の商品のプライシングも行おうというかんじになってきているが。)

3. サーバネットワークのダウンの連鎖

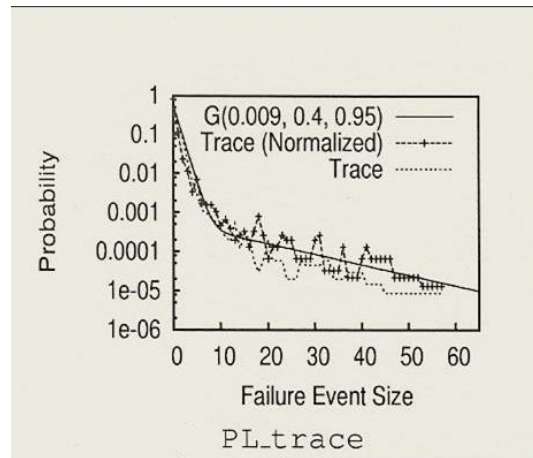


図 6 PlanetLab サーバネットワークにおける同時ダウンの確率分布。横軸は同時ダウン数、縦軸は確率。(文献 [7] より。)

では、インターネットのサーバーにおける連鎖はどうなっているのか？クレジットリスクのマーケットが推定する連鎖倒産の確率分布となにか共通点はあるのだろうか？まず、次の図 6 を見ていただきたい。このグラフは PlanetLab という実証研究のためのサーバネットワークに関するもので、平均 277 台のサーバーの同時ダウンの確率分布を表したものである。確率は、同時ダウン数の増加とともに減少するが、10 台を超えると、減少の割合がゆるやかになっている。また、方対数プロットなのでダウン数がおおきなところでの分布は幾何分布に従っていることがわかる。グラフには、二つの幾何分布を比率 $(1 - \alpha, \alpha)$ の割合で重ね合わせた 2 重幾何分布でフィットした結果もプロットしてある。

$$P_N(n) = G(\alpha, \rho_1, \rho_2) = (1 - \alpha)c_1\rho_1^n + \alpha c_2\rho_2^n$$

$$c_1 = \frac{1 - \rho_1}{1 - \rho_1^{N+1}} \quad c_2 = \frac{1 - \rho_2}{1 - \rho_2^{N+1}} \quad (4)$$

文献には、他にも 2 種類のサーバネットワークに対して、同じ振る舞いを観測し、次の表 3 のような結果にまとめられている。

表 3 サーバネットワークの同時ダウンの確率分布 $G(\alpha, \rho_1, \rho_2)$

Trace	N	α	ρ_1	ρ_2
PL	277	0.009	0.4	0.95
Web	130	0.0012	0.4	0.98
RON	30	0.0012	0.32	0.95

では、なぜサーバーのダウン数の分布が幾何分布に従うのか？また、ダウンの連鎖の強さはどれくらいなのか？もちろん、連鎖倒産のところで行ったのと同じ解析をすれば、確率分布から平均のダウン確率、相関を求めることは容易である。しかし、サーバネットワークにお

いて平均ダウン確率はともかく、平均の相関はあまり意味がない。企業の場合は、資本関係など市場からは不透明な部分が多く、全結合の完全グラフのような依存関係を仮定するしかない場合が多い。また、連鎖の強さもすべてのリンクで同じと仮定していた。この場合、平均相関は意味がある。つまり、情報が不十分である以上、平均相関や条件付相関以上の情報を得ることは難しい。一方、サーバーネットワークの場合は、セキュリティの観点からどのサーバーとどのサーバーが繋がっているのかといった接続情報を隠すことはあるっても、明らかに全結合の完全グラフのようなネットワーク構造というわけではない。

インターネットは、スモールワールド性とスケールフリー性をもったネットワークであることはよく知られ、そうした性質をみたくネットワークの構成方法も数多く提案されている。しかし、複雑ネットワーク上の確率分布の研究は、現在までのところ複雑ネットワーク上の確率モデルの性質に付随して研究される程度であり、一般的な性質はまったく解明されていない。その理由は、確率分布の空間が非常に広大であり、全容解明が難しいからである。たとえば、完全グラフ上の0, 1型の確率変数の場合、変数が交換可能である(区別できない)と仮定すれば、確率分布は変数の数が無限大に拡張できる場合

De Finettiの定理により、2項分布の重ねあわせで表現できる。この場合、相関が負にできないなど、表現できる確率分布は狭い範囲に限定されるが、それでもインジグ模型のような2項分布2個の重ね合わせから、ベータ2項分布のようなベータ関数で滑らかに重ね合わせた場合まで、非常に幅のひろい分布関数を表すことができる。一方、変数の数を無限に伸ばせない場合は、われわれが示した必要条件をみたく条件付確率、相関から確率分布を構成でき、この場合は負の相関をもつ場合まで含めて表現できるが、あくまで必要条件であり、確率分布関数として不適格なものまでも含んでしまう。

このように、完全グラフ上でさえ、確率分布の空間は広大である。そこで、複雑ネット上の0, 1型の変数の確率分布を一般的に構成してその性質を調べるといったことはあきらめ、非常に簡単なネットワーク上での確率変数の振る舞いを調べてみることにする[15]。具体的には次の3つの N サイトからなるネットワーク上の0, 1型の確率変数 $X_i (i = 1, 2, \dots, N)$ で、期待値は $\langle X_i \rangle = p$ に固定するしたものを考える。

- 1次元空間:ただし、サイト上の確率変数は自分のとなりの k サイト上の変数と相関をもつ。 k を $N-1$ に選ぶと、完全グラフとなる。
- 完全グラフ。 $k = N-1$ に対応。
- ケーリー木(ベータ格子)。1個の根から出発し、 z 本の足が次々と生えて成長していくネットワーク。この場合は、最隣接サイト間のみ相関がある($k=1$ の場合)とする。

確率分布は相関のある(つながっている)変数間の条件付確率をあたえることにより構成する。採用する条件付確率は、ベータ2項分布での条件付確率を採用する[10]。

$$\text{Prob}(X_i = 1|n) = p_k(n) = \frac{p(1-\rho) + n\rho}{(1-\rho) + k\rho} \quad (5)$$

この条件付確率を採用したネットワーク上の確率モデルの正体を一言で言えば、「隣接した $k+1$ 個の変数の周辺分布を考えたとき、ベータ2項分布となる模型」である。特に $k=1$ の場合は、隣接する変数が相関 ρ を持つ、非常に単純な模型となる。

同時分布関数を構成してみる。1次元的につながった N 個の確率変数に対し、 $k=1$ の場合は次のようになる。

$$P(x_1, x_2, \dots, x_N) = P(x_1) \prod_{i=1}^{N-1} p(x_{i+1}|x_i)$$

ここで、 $p(x_1) = p^{x_1}(1-p)^{1-x_1}$ 、 $p(x_{i+1}|x_i)$ は、

$$p(x_{i+1}|x_i) = (p_1(x_i)^{x_{i+1}}(1-p_1(x_i))^{1-x_{i+1}})$$

である。一般の k の場合も同様に、 x_2 は x_1 に依存、 x_3 は x_1, x_2 に依存と続け、 $i \geq k+1$ からは、 x_i は、そこまでの k 個の変数に条件付確率(5)でつなげていく。特に、 $k=N-1$ として、すべての確率変数が相関 ρ で結びつく場合、同時分布関数はベータ2項分布のものと同じである。

$$\begin{aligned} P(x_1, x_2, \dots, x_N) \\ = \prod_{i=1}^N (p_{i-1}(\sum_{j=1}^{i-1} x_j))^{x_i} (1 - p_{i-1}(\sum_{j=1}^{i-1} x_j))^{1-x_i} \end{aligned} \quad (6)$$

では、こうして構成した確率モデルはどのような性質を持つだろうか?まず、 x_i と x_{i+j} の相関関数 ρ_j を計算してみる。

$$\rho_j = \frac{\langle X_i X_{i+j} \rangle - p^2}{p(1-p)}$$

$k=1$ の場合は簡単に

$$\rho_j = \rho^j$$

となる。サイト間の距離とともに指数関数的にダンプすることがわかる。一般の k の場合は、次の漸化式が成立する[15]。

$$\rho_{j+k} = \rho \cdot \sum_{i=0}^{k-1} \frac{\rho_{j+i}}{(1-\rho) + k\rho}$$

初期条件は、 $\rho_1 = \rho_2 = \dots = \rho_k = \rho$ である。具体的に ρ_j の j 依存性を書くことは難しいが、ほぼ次のように振る舞う。

$$\rho_{k+j} \propto \rho_{eff}^j \quad j \geq 1$$

ρ_{eff} は次の関係式から求めることができる。

$$\rho_{eff}^k = \frac{\rho}{(1-\rho) + k\rho} \cdot \frac{1 - \rho_{eff}^k}{1 - \rho_{eff}}$$

この結果は、統計物理のスピン模型において、相互作用のレンジをすこし長くしても、レンジが有限であるかぎり、本質的な性質は最隣接相互作用の場合と同じになるという普遍性に対応している。ちなみに、平均相関 ρ_{avg} は次の式で評価できる。

$$\rho_{avg} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \rho_i \cdot (N-i)$$

特に、 $k=1$ の場合は

$$\rho_{avg} = \frac{2}{N(N-1)} \frac{(N-1)\rho - N\rho^2 + \rho^{N+1}}{(1-\rho)^2}$$

N を大きくすると、 ρ_{avg} は $1/N$ のオーダーで小さくなるのがわかる。平均の相関を有限に保つには、 $N \rightarrow \infty$ と $\rho \rightarrow 1$ の二つの極限を同時にとる必要がある。

分布関数はどのように振舞うのか？まず、自明な結論を述べておくと、 p と ρ を固定したまま、 N を大きくすると、平均相関はゼロになり、変数間の相関のない 2 項分布になってしまう。実際のサーバーのダウン数のデータでは、平均相関は有限だったのだから、 N を十分大きくする場合には ρ も 1 に近づけるチューニングを行う必要がある。分布関数 $P_N(n)$ を求めるには、端の変数の値を固定した分布関数を次の式で導入する。

$$\begin{aligned} P_N(n|x_1, x_2, \dots, x_k) \\ \equiv \text{Prob.} \left(\sum_{i=1}^N x_i = n | X_1 = x_1, \dots, X_k = x_k \right) \end{aligned} \quad (7)$$

条件付確率 (5) を用いて、次の漸化式が導かれる。

$$\begin{aligned} P_{N+1}(n|x_1, \dots, x_k) \\ = \sum_{x_{k+1}} P_N(n-1|x_2, \dots, x_{k+1}) p_k(m) x_1 \\ + \sum_{x_{k+1}} P_N(n|x_2, \dots, x_{k+1}) (1-p_k(m))(1-x_k) \\ m = \sum_{i=2}^{k+1} x_i \end{aligned} \quad (8)$$

条件なしの分布関数 $P_N(n)$ を求めるには、条件付の分布関数を条件についてトレースをとればよい。ただし、 $k=1$ の場合には、1 と 0 のドメインウォールが入ると、 $1-\rho$ が現れるので、 $1-\rho$ に関する展開式を求めることができる。具体的な式は複雑だが、 p が小さく、 ρ が 1 に近い場合、近似的に次の式で分布関数が与えられる。

$$P_N(n) \simeq 2(1-\rho) p p_+^{n-1} q q_+^{N-n-1} \simeq \left(\frac{p_+}{q_+} \right)^n$$

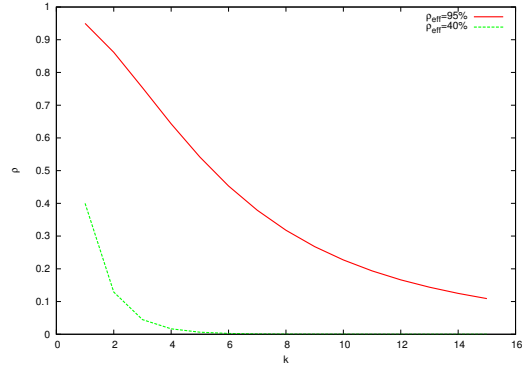


図 7 PL-Trace のパラメータ $\rho_1 = 40\%$ (緑), $\rho_2 = 95\%$ (赤) に対して、ローカルなノード間相関をプロットした。横軸は k で、 $2k+1$ が相互作用のレンジに対応する。

ここで、 $p_+ = p_1(1), q_+ = 1 - p_1(0)$ と定義した。これは幾何分布であり、 p が小さい場合は、

$$P_N(n) \simeq p^n$$

と振舞うことを意味する。一方、 k が 1 よりおおい場合は、 $k=1$ のパラメータ ρ を ρ_{eff} で置き換えた分布とバルクの部分で一致することがわかった。つまり、 p が小さく、相関 ρ が大きい場合は一般的に

$$P_N(n) \simeq \rho_{eff}^n$$

と振舞うことが分かる。サーバーのダウンの同時分布関数から、 ρ_1, ρ_2 というパラメータの幾何分布に従うことが分かっているので、これを ρ_{eff} と読み替えることで、サーバー間の相関を評価することが可能になる。

次の図 7 に、前述の関係式で求めた相関と k の依存性をプロットした。 ρ_{eff} として 95% および 40% を持つ PL-Trace の場合である。 k を大きくするとローカルな相関は減少する。 $\rho_1 = 40\%$ は、 α の値から 99.9% の状況を表しているが、その場合相関は k とともにすぐゼロになる。つまり、ノード間相関はほぼない。一方、0.1% の状況を表している $\rho_2 = 95\%$ では、相関は非常に強く、 $k=5$ でも 50% の相関を示唆している。もちろん、実際のサーバーネットワークは 1 次元ネットワークで依存性のレンジ k を伸ばしたのではなく、スケールフリー性、スモールワールド性をもったものである。そうしたネットワーク上での確率分布の性質を解明していく上での最初の一步となると考えている。

では、この相関の強さはどのような帰結をもたらすのか？そこで、ケーリー木上の $k=1$ の確率模型を用いた相関のあるパーコレーション問題を考える。根 (ノード) が 1 個あり、そこから z 本の枝が生え、ノードを生成し、次からは $z-1$ 本の枝を生成する操作を無限に繰り返して得られるネットワークを考える。そして、ノード上の確率変数の依存関係を $k=1$ の場合の条件付確率 (5) で与えるものとする。興味があるのは、パーコレーション転移の閾値が相関にどう依存するのかである。そのため

には、クラスターサイズの平均値の発散する点を求める必要があり、それは次の対結合関数のノードに関する和で与えられる。

$$G(k) = \langle X_2 X_3 \cdots X_k | X_1 \rangle = p_+^{k-1}$$

クラスターサイズの平均値 $\langle S \rangle$ は、

$$\langle S \rangle = p + \frac{zpp_+}{1 - (z-1)p_+}$$

となり、 $p_+ = \frac{1}{z-1}$ が相転移点であることが分かる。よって、閾値 p_c は、

$$p_c = \left(\frac{1}{z-1} - \rho \right) / (1 - \rho)$$

となり、相関が大きくなると p_c の値は減少し、特に $\rho = \frac{1}{z-1}$ では閾値はゼロ。つまり、たったひとつのノードがダウンすると、そのダウンが端まで伝播することを意味している。

サーバー間のダウンの相関はまれにはあるけれど非常に大きくなり、その場合はたったひとつのサーバーのダウンが系全体の連鎖的なダウンにつながる事が分かる。もちろん、今現在の状態が相関のおおきな状況なのか、小さな状況なのかは判断するのは難しい。けれど、そうしたことが可能なら、相関が高いと判断された場合、サーバーのダウン数がある程度の規模になった場合、極力接続をきって相関を弱めるなどの方法しかシステム全体を守る方法はないのではないかと。

4. これからの課題

以上、企業の連鎖倒産、サーバーの連鎖ダウンの確率分布の振る舞い、そこから連鎖の強さをどのように引き出すのかについて解説してきた。連鎖倒産の場合は、依存関係が不明なので完全グラフを仮定し、クレジットデリバティブのひとつである iTraxx-CJ の6個のプレミアム情報からデフォルト分布関数を推定する。その上で、条件付相関、条件付期待値を求め、クレジットデリバティブ市場で想定されている連鎖の強さ（デフォルト相関）を推定した。またマーケットは中規模な連鎖倒産までしか想定せず、大規模な連鎖倒産は起こらないと考えているらしいこと指摘した。もちろん、エントロピー最大原理を使ってデフォルト分布を推定しているため、推定過程に恣意性がまったく入っていないとはいえない。この点は、Hull らの方法でのインプライド分布でも同様の解析を行い、比較検討する必要があるだろう。

一方、サーバーの連鎖ダウンの確率分布が2種類の幾何分布の重ねあわせとなることについては、簡単なネットワーク上で相関のある確率分布を構成し、ダウン確率が小さく、有限の相関をもつ場合には幾何分布が現れることを示した。そして、サーバー間のダウンの相関がどの程度と見積もれるのかについて、ひとつの見解を述べた。もちろん、サーバーのダウンの確率モデルとして今回

提案したモデルは単純すぎる。1次元格子やケーリー木、完全グラフ上のモデルだけでは不十分であることは間違いない。しかし、ケーリー木上でのパーコレーションの閾値が最隣接ノード間の相関に依存し、相関を大きくすると劇的に小さくなることは、ネットワーク構造がより複雑な場合にも成立する。この結論はネットワークのデザインにおいても重要であり、スケールフリー性がネットワークの頑健さの源ならば、相関は脆弱性の原因である。相関がどの程度ネットワークの頑健さに絡むのかを解明する必要がある。そのためには理論的に取り扱いの容易な一般化ランダムグラフ [1, 2] の場合などに、相関のある2項分布モデルを拡張し、パーコレーション閾値の相関依存性などを解明することが必要である。また、より複雑なネットワーク上の確率分布の振る舞いの研究もこれからの大きな研究課題である。

◇ 参考文献 ◇

- [1] R.Cohen, K.Erez, D. ben-Avraham and S. Havlin: Resilience of the Internet to Random Breakdowns, *Phys.Rev.Lett.***85**4626 (2000).
- [2] D.S. Callaway, M.E.J. Newman, S. H. Strogatz and D.J.Watts : Network Robustness and Fragility : Percolation on Random Graphs , *Phys.Rev.Lett.***85**5468 (2000).
- [3] J.M. Carlson and J. Doyle: Highly optimized tolerance : A mechanism for power laws in designed systems, *Phys.Rev.***E60** 1412 (1999).
- [4] P.J. Schönbucher: Credit Derivatives Pricing Models : Model, Pricing and Implementation, (New York: Wiley).
- [5] M. Bakaloglu, J.Wylie, C.Wang and G. Ganger : On Correlated Failures in Survivable Storage Systems, *Tech. Rep. CMU-CS-02-129*, Carnegie Mellon University (2002).
- [6] P.Yalagandula, S.Nath, H.Yu, P.B. Gibbons and S.Seshan: Beyond Availability :Towards a Deeper Understanding of Machine Failure Characteristics in Large Distributed Systems. In *USENIX WORLDS* (2004).
- [7] S.Nath, H.Yu, P.B.Gibbons and S.Seshan : Subtleties in Tolerating Correlated Failures in Wide-area Storage Systems, *NSDI 3rd Symposium on Networked Systems Design and Implementation* (2006).
- [8] J. Hull and A. White : Credit Derivatives Using an Implied Copula Approach, working paper (University of Toronto).
- [9] S.Mori, K.Kitsukawa and M. Hisakado : Implied default distribution and credit market implications, preprint arXiv:physics0609093.
- [10] M. Hisakado, K. Kitsukawa and S.Mori : Correlated Binomial Models and correlation structures, *J. Phys. A. Math. Gen.* **39**15365 (2006).
- [11] S.Mori, K.Kitsukawa and M. Hisakado : Moody's correlated binomial models for inhomogeneous portfolios, preprint arXiv:physics0603036 .
- [12] K. Kitsukawa, S. Mori and M.Hisakado : Evaluation of Tranche in securitization and long-range Ising model, *Physica A* **368**191 (2006).
- [13] M. Davis and V. Lo ,*Infectious defaults*, *Quantitative Finance*, **1**, 382-387.
- [14] A. Sakata, M. Hisakado and S. Mori : Infectious default model with recovery and continuous limit, *J.Phys.Soc.Jpn* **76** 054801(2007).
- [15] S.Mori and M. Hisakado, in preparation.