

確率的な枝重みをもつ有向非巡回グラフにおける最長路長さの 分布関数の解析的な計算に関する考察

安藤 映 (Ei Ando) 小野 廣隆 (Hirotaka Ono)
定兼 邦彦 (Kunihiko Sadakane) 山下 雅史 (Masafumi Yamashita)

Department of Computer Science and Communication Engineering,
Graduate School of Information Science and Electrical Engineering,
Kyushu University

概要

有向非巡回グラフ (DAG) において、最長路を求める問題は枝重みが確定的な場合には効率的に解けることが知られている。ここで、枝重みが確率変数であるような場合を考えると、最長路長さの分布関数を効率的に求める方法は明らかでない。本稿では、畳み込み積分の数が $O(n)$ 回に抑えることにより、枝重みの分布の形態によっては DAG の最長路長さの分布関数の解析的な表現が効率的に求まり得ることを示す。特に、枝重みの従う分布としてすべて同一のパラメータをもつ指数分布を仮定した場合、最長路長さの分布関数の記述は $O(mn)$ の長さで初等関数の組み合わせとして表現できることを示す。ただし、 n はグラフの頂点の数、 m はグラフの枝の数である。

1 はじめに

有向非巡回グラフ (DAG) の最長路問題を考える。この問題は枝重みが確定的な場合には効率的に解くことが可能である。しかし、枝重みが確率変数であるような場合には次の二つのアプローチを別々に考える必要がある。(1) 最長路になる確率をもっとも高い路を一つ見つける (2) DAG の中の最長路長さを一つの確率変数と捉え、その分布を求める。これらのどちらも確定的な枝重みの場合よりも大きく難しくなることが知られているが、本稿では (2) の立場をとる。

確率的な枝重みをもつ DAG 最長路問題において、その分布を求めることは論理回路の信号遅延を求める観点から活発に近似アルゴリズムの研究がなされている [1, 2, 3, 4]。ただし、これらの近似手法は見積りに関して近似保証が与えられていないため、近似保証のある手法が [5] において提案されている。

近似ではなく、厳密な分布を求めようとする研究もあり、Kulkarni [6] は最小全域木問題において互いに独立で、指数分布に従う確率変数の枝重みを仮定して、任意の精度でその分布関数を求める手法を提案している。また、Martin [7] は DAG の最長路問題において、枝重みを多項式分布に従う互いに独立な確率変数とし、全ての路の列挙を行った上で最長路長さ分布を求めることを提案している。ただし、これら二つの手法はどちらもグラフ規模の指数に比例する計算時間を必要とする。

本稿では、畳み込み積分の数を $n-1$ 回に抑えることにより、枝重みの分布の形態によっては DAG の最長路長さの分布関数の解析的な表現が効率的に求まり得ることを示す。次に、全ての枝重みが互いに独立でかつ同一のパラメータの指数分布に従う場合、その分布関数の式は $O(mn)$ 個の項の和として数式表現できることを示す。ただし、 n はグラフの頂点の数、 m はグラフの枝の数である。

2 準備

頂点集合 $V = \{v_1, v_2, \dots, v_n\}$ と有向枝集合 $E \subset V \times V$ が与えられ、グラフ $G = (V, E)$ の中に閉路が無いとき、 G は有向非巡回グラフ (DAG) であると言う。入次数が 0 の頂点と出次数が 0 の頂点をそれぞれ G の入口、出口と言う。枝集合 E のそれぞれの要素 (v_i, v_j) に確率変数 X_{ij} の重みがあるとする。グラフ G 中の入口から出口までの路全体の集合を Π とすると、 G 中の最長路の長さ X は確率変数として次の式で表され、本研究ではこの確率変数の分布を求める。

$$X = \max_{\pi \in \Pi} \left\{ \sum_{\{v_i, v_j\} \in \pi} X_{ij} \right\}$$

確率変数 X について、その分布関数 $F(x) = P(X \leq x)$ が次の式で表されるとき、確率変数 X は指数分布に従うと言う。

$$F(x) = \begin{cases} 1 - \exp(-a(x - b)) & x \geq 0 \\ 0 & x < 0 \end{cases}$$

また、確率分布関数の微分 $f(x) = F'(x)$ を確率密度関数と言う。

二つの互いに独立な確率変数 X_1, X_2 の確率分布関数をそれぞれ $F_1(x), F_2(x)$ 、確率密度関数をそれぞれ $f_1(x), f_2(x)$ とする。二つの確率変数の和 $X_1 + X_2$ の確率分布関数は次の計算により、片方の確率分布関数 $F_1(x)$ ともう一方の確率密度関数 $f_2(x)$ の畳み込み積分として得られることが言える。

$$\begin{aligned} P(X_1 + X_2 \leq x) &= \int_{\mathbf{R}} P(X_1 + t \leq x \mid X_2 = t) f_2(t) dt \\ &= \int_{\mathbf{R}} P(X_1 \leq x - t \mid X_2 = t) f_2(t) dt \\ &= \int_{\mathbf{R}} P(X_1 \leq x - t) f_2(t) dt \\ &= \int_{\mathbf{R}} F_1(x - t) f_2(t) dt \end{aligned}$$

また、二つの確率変数の最大値 $\max\{X_1, X_2\}$ の確率分布関数は次の計算によって $F_1(x)$ と $F_2(x)$ の積として得られることが言える。

$$P(\max\{X_1, X_2\} \leq x) = P(X_1 \leq x \wedge X_2 \leq x) = F_1(x)F_2(x)$$

3 $n - 1$ 回の畳み込み積分による最長路長さ分布の計算

本節では DAG の最長路長さの分布関数が $n - 1$ 回の畳み込み積分によって求められることを示す。求める DAG の最長路長さの分布関数は、次の形で表現できる。

$$X = P \left(\bigwedge_{\pi \in \Pi} \left(\sum_{e \in \pi} X_e \leq x \right) \right) \quad (1)$$

式 (1) の形は記述の上では簡潔であるが、この式は全ての路の列挙が必要で、かつ DAG の路の数は最大で 2^{n-2} になり得るため、このままでは効率的な計算へ進む見込みが立てられない。しかし、次の定理によって式 (1) を $n - 1$ 回の畳み込み積分で表現できることが言える。

定理 1 枝 (v_i, v_j) の重み X_{ij} の分布関数を $F_{ij}(x)$ とし、もしも $(v_i, v_j) \notin E$ の場合には任意の x について $F_{ij}(x) = 1$ とする。入口と出口を一つずつ持つ DAG G の入口から出口までの路全体の集合を Π とすると、最長路長さの分布関数は次の式で表される。

$$P\left(\bigwedge_{\pi \in \Pi} \left(\sum_{e \in \pi} X_e \leq x\right)\right) = \int_{\mathbf{R}^{n-1}} H(x - z_1) \prod_{1 \leq i \leq n-1} \left(\frac{d}{dz_i} \prod_{i+1 \leq j \leq n} F_{ij}(z_i - z_j)\right) dz_i$$

ただし、 $H(x)$ は次の式で定義される関数 (ステップ関数) である。

$$H(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

証明 与えられた任意の DAG $G = (V, E)$ の最長路長さの分布を求めるために、各頂点 v_i から $i < j$ を満たす任意の頂点 v_j に対して枝があるような n 頂点の確率変数重み付きグラフ $\vec{K}_n$ を考える。つまり、 $\vec{K}_n = (V, E_K)$ とすると $E_K = \{(v_i, v_j) \mid 1 \leq i < j \leq n\}$ である。グラフ $\vec{K}_n$ の枝重みは、もし $(v_i, v_j) \in E$ の場合、グラフ G の中で枝 (v_i, v_j) の枝重みを用いる。もし $(v_i, v_j) \notin E$ の場合、枝重み X_{ij} の分布関数 $P(X_{ij} \leq x)$ を全ての x について $F_{ij}(x) = 1$ とする。こうすることにより、 $(v_i, v_j) \notin E$ の場合には X_{ij} は確率 1 で十分小さな値を取る確率変数であるとみなすことができる。最長路問題を考える場合、そのような重みをもつ枝 (v_i, v_j) は最長路に含まれることがないため、 $\vec{K}_n$ に (v_i, v_j) が含まれていない場合と等価であり、 $\vec{K}_n$ の最長路長さの分布は G の最長路長さの分布と等しい。従って、以下では分布関数の計算を $\vec{K}_n$ に対して行う。

頂点 v_i から v_j までの全ての路の集合を $\Pi(i, j)$ とする。すると、式 (1) は次のように書き直すことができる。

$$P\left(\bigwedge_{\pi \in \Pi(1, n)} \left(\sum_{(i, j) \in \pi} X_{ij} \leq x\right)\right) \quad (2)$$

そして、各頂点 v_i から出口 v_n までの最長路長さを表す変数を $Z_i = \max_{\pi \in \Pi(i, n)} \{\sum_{(i, j) \in \pi} X_{ij}\}$ とする。すると、 $Z_{n-1} = X_{n-1, n}$ であって、他のどの X_{ij} とも独立であるので、次のように式 (2) の変形を行うことができる。

$$P\left(\bigwedge_{\pi \in \Pi_{n-1}(1, n)} \left(\sum_{(i, j) \in \pi} X_{ij} \leq x\right) \wedge \bigwedge_{\pi \in \Pi(1, n-1)} \left(\sum_{(i, j) \in \pi} X_{ij} + Z_{n-1} \leq x\right)\right) \quad (3)$$

ただし、 $\Pi_k(i, j)$ は v_i から v_j までの路のうち、 $U_k = \{v_k, v_{k+1}, \dots, v_{n-1}\}$ を通らない路の集合とする。確率変数 Z_{n-1} の分布関数を $G_{n-1}(x)$ と置くと、式 (3) は次のように変形できる。

$$\int_{\mathbf{R}} P\left(\bigwedge_{\pi \in \Pi_{n-1}(1, n)} \left(\sum_{(i, j) \in \pi} X_{ij} \leq x\right) \wedge \bigwedge_{\pi \in \Pi(1, n-1)} \left(\sum_{(i, j) \in \pi} X_{ij} + z_{n-1} \leq x\right) \middle| Z_{n-1} = z_{n-1}\right) dG_{n-1}(z_{n-1}) \quad (4)$$

この式 (4) は $Z_{n-1} = X_{n-1, n}$ が他のどの枝重みとも互いに独立であるので、次のように変形できる。

$$\int_{\mathbf{R}} P\left(\bigwedge_{\pi \in \Pi_{n-1}(1, n)} \left(\sum_{(i, j) \in \pi} X_{ij} \leq x\right) \wedge \bigwedge_{\pi \in \Pi(1, n-1)} \left(\sum_{(i, j) \in \pi} X_{ij} + z_{n-1} \leq x\right)\right) dG_{n-1}(z_{n-1}) \quad (5)$$

定義により, $G_{n-1}(z_{n-1}) = P(X_{n-1,n} \leq x)$ なので, 式 (5) は次のように変形できる.

$$\int_{\mathbf{R}} P \left(\bigwedge_{\pi \in \Pi_{n-1}(1,n)} \left(\sum_{(i,j) \in \pi} X_{ij} \leq x \right) \wedge \bigwedge_{\pi \in \Pi(1,n-1)} \left(\sum_{(i,j) \in \pi} X_{ij} + z_{n-1} \leq x \right) \right) \left(\frac{d}{dz_{n-1}} P(X_{n-1,n} \leq z_{n-1}) \right) dz_{n-1} \quad (6)$$

式 (2) から式 (6) への変形を同様にして $Z_{n-2}, Z_{n-3}, \dots, Z_2$ の順で行う. 式変形の繰り返しのにより, 次の式を得ることができる.

$$\int_{\mathbf{R}^{n-2}} P \left(\bigwedge_{2 \leq l \leq n} \bigwedge_{\pi \in \Pi_2(1,l)} \left(\sum_{(i,j) \in \pi} X_{ij} + z_l \leq x \right) \wedge \bigwedge_{\pi \in \Pi(1,2)} \left(\sum_{(i,j) \in \pi} X_{ij} + z_2 \leq x \right) \right) \prod_{1 \leq i \leq n-1} \left(\frac{d}{dz_i} \prod_{i+1 \leq j \leq n} P(X_{ij} + z_j \leq z_i) \right) dz_i \quad (7)$$

ただし, 式 (7) では $z_n = 0$ とする.

定義より, $\Pi(1,2)$ は一本の枝 (v_1, v_2) しか含まず, また $\Pi_2(1,l)$ は常に一本の枝 (v_1, v_l) のみしか含まないので, 式 (7) は次のように変形できる.

$$\int_{\mathbf{R}^{n-2}} P \left(\bigwedge_{2 \leq l \leq n} (X_{1l} + z_l \leq x) \right) \prod_{2 \leq i \leq n-1} \left(\frac{d}{dz_i} \prod_{i+1 \leq j \leq n} P(X_{ij} + z_j \leq z_i) \right) dz_i \quad (8)$$

ここで $F_{ij}(x) = P(X_{ij} \leq x)$ とおくと, 次の式が得られる.

$$\int_{\mathbf{R}^{n-2}} \prod_{2 \leq l \leq n} F_{1,l}(x - z_l) \prod_{2 \leq i \leq n-1} \left(\frac{d}{dy_i} \prod_{i+1 \leq j \leq n} F_{ij}(z_i - z_j) \right) dz_i \quad (9)$$

この式をステップ関数 $H(x)$ を用いて整理すると定理が得られる. □

入口や出口が複数あるような DAG の場合も元のグラフ G グラフに確定的な値 0 を重みとする枝を追加したグラフ G' を考えることで対応できる. まず, 複数ある入口のうちどれか一つを v_1 とし, v_1 から残りの入口への重み 0 の枝を追加する. 次に, 複数ある出口のうち一つを v_n とし, 残りの出口それぞれから v_n への重み 0 の枝を追加する. こうしてできたグラフを G' とすると, 重み 0 の枝重みの確率分布関数はステップ関数 $H(x)$ で与えられるため定理 1 によって計算できる.

定理 1 で与えられた式は煩雑で, すぐに意味を読み取ることは難しいかもしれない. しかし, 枝重みが確定的な値の場合との対比をすることで理解が助けられる. それぞれの枝 $(v_i, v_j) \in E$ の重みがある確定的な値 c_{ij} の場合, この枝重みは分布関数 $H(x - c_{ij})$ をもつ確率変数とみなすことができる. このとき, 定理 1 の式中で分布関数の積は, 各頂点の枝重みの最大値を取る演算と対応し, また, z_i に関する畳み込み積分は頂点 v_i から出口までの最長路長さを現在の計算結果に加算することに対応する. 従って, 定理 1 は枝重みが確定的な値の場合にも適用でき, 定理 1 が確定的な場合の最長路問題の解法の自然な拡張として得られたことがわかる.

4 枝重みが指数分布に従う場合の最長路長さの分布

前節の定理により, 枝重みの分布関数が積分計算のし易い形で与えられている場合には DAG の最長路長さの分布関数も解析的に計算を行うことが可能である. 本節ではまず枝重みが指数分布に従う確率変数の場合に, 最長路長さが初等関数の組み合わせとして解析的に得られることを示す.

定理 2 枝重みが互いに独立かつ指数分布に従う確率変数または確定的な値として与えられている DAG を G とする。するとグラフ G の最長路長さの分布関数は初等関数とステップ関数の組み合わせで表現できる。

証明 指数分布の分布関数やステップ関数の畳み込み積分を計算した場合に得られる項は、ステップ関数、 x の多項式、 x の指数関数、定数またはそれらの組み合わせの積であり、そのような関数を多項式指数分布関数と呼ぶことにする。すると、ステップ関数、 x の多項式、 x の指数関数、定数のうちの二つを互いに畳み込み積分しても多項式指数分布関数であるため、多項式指数分布関数の間の畳み込み積分は多項式指数分布関数である。□

次に、枝重みの従う分布としてパラメータがすべて同一の指数分布を仮定した場合、最長路長さの分布関数の記述の長さ $O(mn)$ であることを示す。

定理 3 全ての枝重みが次の指数分布関数 $F(x)$ で与えられる確率変数であるような DAG の最長路問題を考える。

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-x) & x \geq 0 \end{cases} \quad (10)$$

このとき、最長路長さの分布関数をステップ関数及び初等関数の組み合わせとして表現すると、 $O(mn)$ 個の項の和にまとめることができる。

証明 証明は全ての枝重みが 0 または分布関数 $F(x)$ を持つ確率変数であるような DAG について定理 1 の計算を行い、 $O(mn)$ 個の項の和にまとめられることを帰納法で示す。

定理 1 において、まず z_1 に関する畳み込み積分をすると、式 (9) において $(v_i, v_j) \in E$ の場合に $F_{ij}(x) = F(x)$ とした式が得られる。ここで、仮に $v_2, \dots, v_{n-1}$ から出口までの最長路の長さが全て確定的な値 0 であると置き換えをすれば、式 (9) はある自然数 k について $(F(x))^k$ として与えられ、この式は $\exp(-lx)$ の和として与えられる。次に、 $v_2, \dots, v_{n-1}$ から出口までの最長路の長さが 0 でない場合を考える。今、入口から z_i までの積分が完了し、 $v_{i+1}, \dots, v_{n-1}$ から出口 v_n までの最長路の長さを確定的な値 0 と置き換えた場合には次の式 (11) で表される関数 $G_i(x)$ として最長路長さの分布関数が表現できると仮定する。

$$G_i(x) = \sum_{1 \leq a \leq a_i - n \leq b \leq b_i} C_{ab} x^a \exp(-bx) \quad (11)$$

ただし C_{ab} は (a, b) の組み合わせ毎に定義される定数とする。ここで、 v_{i+1} から出る枝の重みのうち h 本は分布関数 $F(x)$ をもつ確率変数であり、 $v_{i+2}, \dots, v_{n-1}$ から出口までの最長路長さを 0 で置き換えた場合を考える。すると、この場合に z_{i+1} に関する畳み込み積分を行った結果 $G_{i+1}(x)$ は次式の形で表現できる。

$$G_{i+1}(x) = \sum_{1 \leq a \leq a_{i+1} - n \leq b \leq b_{i+1}} D_{ab} x^a \exp(-bx) \quad (12)$$

ただし、 D_{ab} は (a, b) の組み合わせ毎に定義されるある定数である。式 (12) 中の a_{i+1} と b_{i+1} の大きさを評価する。直接 $G_i(x)$ から $G_{i+1}(x)$ を計算することはできないが、 $G_i(x)$ は、定理 1 に基づいて $z_1, \dots, z_{i-1}$ に関する畳み込み積分を計算した関数 $L_i(x, z_i, \dots, z_{n-1})$ において $z_i = z_{i+1} = \dots = z_{n-1} = 0$ として得られる関数であり、 $G_{i+1}(x)$ は $L_{i+1}(x, z_{i+1}, \dots, z_{n-1})$ について $z_{i+1} = \dots = z_{n-1} = 0$ として得られる関数である。ここで L_i から L_{i+1} を求める際には z_i に関する畳み込み積分を行うが、 L_i の式中には z_i に関わらない x を含む項も存在する。従って、次の式 (13) の x の次数を a'_{i+1} 及び $\exp$ 中の x の係数を b'_{i+1} とすると、 $a'_{i+1} \geq a_{i+1}$ 及び $b'_{i+1} \geq b_{i+1}$ が成立する。

$$\int_0^x G_i(x - z_{i+1}) \frac{d}{dx} (F(z_{i+1}))^h dz_i \quad (13)$$

畳み込み積分の結果には x^{a_i+1} よりも x の次数が高い項が現れないので $a'_{i+1} - a_i$ は高々1である。また、積分によって指数の中の x の係数は変化しないので $b'_{i+1} - b_i$ は高々 h である。

以上により、DAG の最長路長さの分布関数 $G_{n-1}(x)$ は式 (11) において $i = n-1$ とした形で表され、さらに $a_{n-1} \leq n-2$ かつ $b_{n-1} \leq m$ ($\because h \leq \deg(v_i)$) が言える。□

5 おわりに

本研究では、DAG の最長路問題において、枝重みが互いに独立な場合に最長路長さの分布関数が頂点数 n に比例した数の畳み込み積分と微分の組み合わせによって表現されることを示し、枝重みが指数分布に従う場合について一般の DAG について、最長路長さを $O(n)$ 回の畳み込み積分によって初等関数を用いた表現ができることを示した。これにより、全ての枝重みが互いに独立かつパラメータが同一の指数分布に従う場合には、最長路長さの分布関数は初等関数の組み合わせとして $O(mn)$ の長さの数式で表現できる。本研究で得られた式を用いれば、枝重みとして他の確率分布 (一般の指数分布、一様分布や多項式分布) を仮定しても最長路長さの分布を解析的に計算することが可能である。ただし、畳み込み積分が $O(n)$ 回に抑えられても解析的な分布を表現する数式や途中計算の式が非常に長なる可能性があり、このため本研究の手法においてもまだ効率的な手法が得られたわけではない。

参考文献

- [1] M.Berkelaar, Statistical delay calculation, a linear time method. *Proceedings of the International Workshop on Timing Analysis (TAU'97)*, pp.15-24, 1997.
- [2] M.Hashimoto and H.Onodera, A performance optimization method by gate sizing using statistical static timing analysis. *IEICE Trans. Fundamentals*, E83-A, 12, pp.2558-2568, 2000.
- [3] 築山 修治, 田中 正和, 福井 正博, 組み合わせ回路におけるクリティカルパス遅延のばらつき見積り手法, 第13回回路とシステムワークショップ, pp.131-135, 2000.
- [4] E. Ando, M. Yamashita, T. Nakata, Y.Matsunaga, The Statistical Longest Path Problem and Its Application to Delay Analysis of Logical Circuits, *proc.TAU*, 2002, pp.134-139.
- [5] E.Ando, T.Nakata, M.Yamashita, Approximating the Longest Path Length of a Stochastic DAG by a Normal Distribution in Linear Time. *Submitted for publication*.
- [6] V.G.Kulkarni, Minimal Spanning Trees in Undirected Networks with Exponentially Distributed Arc Weights, *NETWORKS*, Vol. 18,1988, pp.111-124.
- [7] J.J.Martin, Distribution of the time through a directed, acyclic network, *Operations Research* 13 (1965) 46-66.