

Weight of the Shortest Path to the First Encountered Peer in a Peer Group of Size m

P. Van Mieghem* and S. Tang†

Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands

June 26, 2007

Abstract

We model the weight (e.g. delay, distance or cost) from an arbitrary node to the nearest (in weight) peer in a peer-to-peer (P2P) network. The exact probability generating function and an asymptotic analysis is presented for a random graph with i.i.d. exponential link weights. The asymptotic distribution function is a Fermi-Dirac distribution that frequently appears in statistical physics. The good agreement with simulation results for relatively small P2P networks makes the asymptotic formula for the probability density function useful to estimate the minimal number of peers to offer an acceptable quality (delay or latency).

1 Introduction

The present analysis is motivated by content distribution in peer-to-peer (P2P) networks (see e.g. [6, Chapter 13]) such as Napster, Bit Torrent, Tribler etc. Content is often either fully replicated at a peer or is split into chunks and stored over m peers. A major task of a member of the peer group lies in selecting the best peer among those m peers that possess the desired content. This searching process for peers is not always optimized and several variants exist depending on which criterion is optimized. For example, one may choose that peer that is most nearby in the number of hops or in the delay or latency. The first problem is analogous to the anycast problem in IPv6, treated in [5, Chapter 18]. Here, we confine to the second problem: given a network of N nodes over which m peers are randomly distributed, what is the delay from a certain peer to the nearest (in delay) peer? The presented analysis shows how that delay varies as the number m of peers changes and it may set bounds on the minimum number of peers needed to still offer an acceptable content distribution service.

In a P2P network, peers are joining and leaving regularly which motivates us to consider such a network to first order as an Erdős-Rényi random graph. The shortest path between two nodes is the path for which the sum of the link weights is minimal. The shortest path tree (SPT) rooted at an arbitrary node to m uniformly chosen peers is the union of the shortest paths from that node to all m different peers in the graph. Thus, the computation of a shortest path requires the knowledge of link weights, such as a monetary cost, the number of hops, a delay, a distance, etc. In most communication networks, link weights are not precisely known. Here and as motivated in [5, Section 16.1], we assign i.i.d. exponentially distributed weights with unit mean to links.

In [5, 4], we have rephrased the shortest path problem between two arbitrary nodes in the complete graph K_N with i.i.d. exponential link weights as a Markov discovery process, that starts the path searching process at the source. The discovery process is a continuous time Markov chain with N states. Each state n represents the n already discovered nodes (including the source node). If at some stage in the Markov discovery process n nodes are discovered, then the next node is reached with rate $\lambda_{n,n+1} = n(N - n)$, which is the transition rate

*Email: P.F.A.VanMieghem@tudelft.nl.

†Email: S.Tang@ewi.tudelft.nl.

in the continuous time Markov chain. Since the discovering of nodes at each stage only increases n , the Markov discovery process is a pure birth process with birth rate $n(N - n)$. We call τ_n the inter-attachement time between the inclusion of the n -th and $(n + 1)$ -th node to the SPT for $n = 1, \dots, N - 1$. The inter-attachement time τ_n is exponentially distributed with parameter $n(N - n)$ as follows from the theory of Markov processes. By the memoryless property of the exponential distribution and the symmetry in the complete graph K_N , the new node is added uniformly to an already discovered node. Hence, the resulting SPT to all nodes, i.e. $m = N - 1$, is exactly a uniform recursive tree (URT). A URT of size N is a random tree rooted at some source node and where at each stage a new node is attached uniformly to one of the existing nodes until the total number of nodes is equal to N . As proved in [4] for large N , the URT is also asymptotically the shortest path in the class of Erdős-Rényi random graph with i.i.d. exponential link weights.

The paper is outlined as follows. Section 2 derives the exact probability generating function of the weight towards the first encountered peer. The asymptotics for large N (and constant peer group size m) is computed in Section 3 and compared with simulations for finite graph size N . Mathematical derivations are given in the appendices.

2 The probability generating function

The weight $W_{N;m}$ of the shortest path from an arbitrary node to the nearest (in weight) peer in a peer group of size m is¹

$$W_{N;m} = \sum_{k=1}^{N-1} 1_{\{T_N(m) \geq k\}} \tau_k \quad (1)$$

where $T_N(m)$ is the number of steps in the continuous Markov discovery process until one peer of the group of m peers is reached, or, $T_N(m)$ is the hitting time to the peer group of size m . Since the random variables $1_{\{T_N(m) \geq k\}}$ and τ_k (for a fixed k) are independent, the average weight directly follows as

$$E[W_{N;m}] = \sum_{k=1}^{N-1} \Pr[T_N(m) \geq k] E[\tau_k]$$

The event $\{T_N(m) \geq k\}$ implies that the k -th discovered (attached) node in the URT is the first peer out of the m peers that is reached. Hence, the $k - 1$ previously discovered nodes do not belong to the peer group. Since the m peers as well as the $k - 1$ nodes are uniformly chosen out of the $N - 1$ nodes, we have that

$$\Pr[T_N(m) \geq k] = \frac{\binom{N-1-m}{k-1}}{\binom{N-1}{k-1}}$$

such that, with $E[\tau_k] = \frac{1}{k(N-k)}$,

$$E[W_{N;m}] = \frac{(N-1-m)!}{(N-1)!} \sum_{k=1}^{N-1} \frac{(N-1-k)!}{k(N-m-k)!}$$

Invoking the identity (15) yields

$$\sum_{k=1}^{N-m} \frac{1}{k} \frac{(N-1-k)!}{(N-m-k)!} = \frac{(N-1)!}{(N-m)!} \sum_{j=1}^{N-m} \frac{1}{j} - \frac{(N-1)!}{(N-m-1)!} \sum_{q=0}^{m-2} \frac{1}{(m-1-q)(N-1-q)}$$

After partial fraction expansion, the q -sum is

$$\sum_{q=0}^{m-2} \frac{1}{(m-1-q)(N-1-q)} = \frac{1}{N-m} \sum_{j=1}^{m-1} \frac{1}{j} - \frac{1}{N-m} \sum_{j=N-m+1}^{N-1} \frac{1}{j}$$

¹This has been conveyed to us by Professor S. M. Ross.

and we arrive at

$$E[W_{N;m}] = \frac{1}{N-m} \left\{ \sum_{j=1}^{N-1} \frac{1}{j} - \sum_{j=1}^{m-1} \frac{1}{j} \right\} = \frac{\psi(N) - \psi(m)}{N-m} \quad (2)$$

where $\psi(x)$ is the digamma function [1, Section 6.3]. For large N , we have [1, Section 6.3.18]

$$E[W_{N;m}] = \frac{\ln \frac{N}{m}}{N-m} + \frac{1}{2Nm} + O(N^{-1}m^{-2})$$

Since the sequence of random variables $1_{\{T_N(m) \geq 1\}}, 1_{\{T_N(m) \geq 2\}}, \dots, 1_{\{T_N(m) \geq N-1\}}$ is obviously not independent, a computation of the probability generating function $\varphi_{W_{N;m}}(z) = E[e^{-zW_{N;m}}]$ different from straightforwardly using (1) is needed. We define $v_k = \sum_{n=1}^k \tau_n$ as the weight of the shortest path to the k -th attached (discovered) node. Since the laws of the Markov discovery process and of the URT are independent [4], we obtain

$$\varphi_{W_{N;m}}(z) = \sum_{k=1}^{N-m} E[e^{-zv_k}] \Pr[Y_m(k)]$$

where $Y_m(k)$ denotes the event that the k -th attached node is the first encountered peer among the m peers in the URT. Now, there are $\binom{N-1}{m}$ ways to distribute the m peers (different from the source node) over the $N-1$ remaining nodes. If the k -th node is the first of the m discovered peers, it implies that the remaining $m-1$ peers need to be discovered later. There are $\binom{N-1-k}{m-1}$ possible ways, whence

$$\Pr[Y_m(k)] = \frac{\binom{N-1-k}{m-1}}{\binom{N-1}{m}}$$

Thus, we obtain²

$$\varphi_{W_{N;m}}(z) = \frac{m(N-1-m)!}{(N-1)!} \sum_{k=1}^{N-m} \frac{(N-1-k)!}{(N-m-k)!} \prod_{n=1}^k \frac{n(N-n)}{z+n(N-n)} \quad (3)$$

After partial integration of a single-sided Laplace transform $\varphi_X(z) = \int_0^\infty f_X(t) e^{-zt} dt$ and assuming that the derivative $f'_X(t)$ exists, the relation $f_X(0) = \lim_{z \rightarrow \infty} z\varphi_X(z)$ is found. Applied to (3) yields

$$f_{W_{N;m}}(0) = m \quad (4)$$

3 The asymptotic pgf and pdf

Although the inverse Laplace transform of (3) can be computed, the resulting series for the pdf is hardly appealing. As shown below, an asymptotic analysis leads to an elegant result that, in addition, seems applicable to a relatively small network size N . Following a similar procedure as [5, pp. 518-520], we write

$$z + n(N-n) = \left(\sqrt{\left(\frac{N}{2}\right)^2 + z} + \frac{N}{2} - n \right) \left(\sqrt{\left(\frac{N}{2}\right)^2 + z} - \left(\frac{N}{2} - n\right) \right)$$

and define $y = \sqrt{\left(\frac{N}{2}\right)^2 + z}$. Then,

$$\begin{aligned} \prod_{n=1}^k \frac{n(N-n)}{z+n(N-n)} &= \frac{k!(N-1)!}{(N-k-1)!} \prod_{n=1}^k \frac{1}{\left(y + \frac{N}{2} - n\right)} \prod_{n=1}^k \frac{1}{\left(y - \frac{N}{2} + n\right)} \\ &= \frac{k!(N-1)!}{(N-k-1)!} \frac{\Gamma\left(y + \frac{N}{2} - k\right)}{\Gamma\left(y + \frac{N}{2}\right)} \frac{\Gamma\left(y - \frac{N}{2} + 1\right)}{\Gamma\left(y - \frac{N}{2} + k + 1\right)} \end{aligned}$$

²It is readily verified that by computing $E[W_{N;m}] = - \left. \frac{d\varphi_{W_{N;m}}(z)}{dz} \right|_{z=0}$ the relation (2) is obtained.

and, substituted in (3), yields

$$\varphi_{W_{N;m}}(z) = \frac{m(N-1-m)! \Gamma\left(y - \frac{N}{2} + 1\right)}{\Gamma\left(y + \frac{N}{2}\right)} \sum_{k=1}^{N-m} \frac{k!}{(N-m-k)!} \frac{\Gamma\left(y + \frac{N}{2} - k\right)}{\Gamma\left(y - \frac{N}{2} + k + 1\right)}$$

For large N and $|z| < N$, we have³ that $y = \sqrt{\left(\frac{N}{2}\right)^2 + z} \sim \frac{N}{2} + \frac{z}{N}$ such that

$$\varphi_{W_{N;m}}(z) \sim \frac{m(N-1-m)! \Gamma\left(\frac{z}{N} + 1\right)}{\Gamma\left(N + \frac{z}{N}\right)} \sum_{k=1}^{N-m} \frac{k!}{(N-m-k)!} \frac{\Gamma\left(N + \frac{z}{N} - k\right)}{\Gamma\left(\frac{z}{N} + k + 1\right)}$$

We now introduce the scaling $z = Nx$, where $|x| < 1$ since $|z| < N$,

$$\varphi_{W_{N;m}}(Nx) \sim \frac{m(N-1-m)! \Gamma(x+1)}{\Gamma(N+x)} \sum_{k=1}^{N-m} \frac{k!}{(N-m-k)!} \frac{\Gamma(N+x-k)}{\Gamma(x+1+k)} \quad (5)$$

For large N and fixed m and x , the asymptotic order of the sum

$$S = \sum_{k=1}^{N-m} \frac{\Gamma(k+1)}{\Gamma(k+x+1)} \frac{\Gamma(N-k+x)}{\Gamma(N-k-m+1)} \quad (6)$$

scales as

$$S = \frac{\Gamma(N+x+1) N^{-x-1}}{\Gamma(x) \Gamma(N-m)} \frac{\pi}{\sin \pi x} \frac{\Gamma(x+m)}{m!} (1 + O(N^{-1})) \quad (7)$$

This result (7) is derived in Section A. Substitution of (7) into (5), leads, for large N , fixed m and $|x| < 1$, to

$$\varphi_{W_{N;m}}(Nx) \sim \frac{\pi x}{\sin \pi x} \frac{\Gamma(x+m)}{\Gamma(m)} N^{-x} (1 + O(N^{-1}))$$

or

$$\lim_{N \rightarrow \infty} N^x \varphi_{W_{N;m}}(Nx) = \lim_{N \rightarrow \infty} E \left[e^{-(NW_{N;m} - \ln N)x} \right] = \frac{\pi x}{\sin \pi x} \frac{\Gamma(x+m)}{\Gamma(m)} \quad (8)$$

Finally, the inverse Laplace transform, computed in Section B, is

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln N \leq t] = m e^{-mt} e^{e^{-t}} \int_{e^{-t}}^{\infty} \frac{e^{-u}}{u^{m+1}} du = m e^{-mt} e^{e^{-t}} \Gamma(-m, e^{-t}) \quad (9)$$

where $\Gamma(-m, e^{-t})$ is the incomplete Gamma function (17). If $m = 1$, then we have

$$\lim_{N \rightarrow \infty} \Pr[NW_N - \ln N \leq t] = 1 - e^{-t} e^{e^{-t}} \int_{e^{-t}}^{\infty} \frac{e^{-u}}{u} du \quad (10)$$

The probability density function follows after derivation of (9) with respect to t as

$$\lim_{N \rightarrow \infty} f_{NW_{N;m} - \ln N}(t) = m - m e^{-mt} (e^{-t} + m) e^{e^{-t}} \int_{e^{-t}}^{\infty} \frac{e^{-u}}{u^{m+1}} du \quad (11)$$

3.1 The Fermi-Dirac distribution

For large m and to first order, we have $\frac{\Gamma(x+m)}{\Gamma(m)} \sim m^x$, such that

$$\lim_{N \rightarrow \infty} E \left[e^{-(NW_{N;m} - \ln N/m)x} \right] = \frac{\pi x}{\sin \pi x}$$

In view of the average (2) for large N , we can write $NW_{N;m} - \ln N/m = N(W_{N;m} - E[W_{N;m}])$. After inverse Laplace transform with $0 < c < 1$,

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln N/m \leq t] = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\pi e^{tx}}{\sin \pi x} dx$$

³The notation $f(x) \sim g(x)$ for large x means that $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1$.

For $t > 0$ ($t < 0$), the contour can be closed over the negative (positive) $\text{Re}(x)$ -plane, resulting, for any t , in the Fermi-Dirac distribution function

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln N/m \leq t] = \frac{1}{1 + e^{-t}} \quad (12)$$

whose symmetric probability density function is

$$\lim_{N \rightarrow \infty} f_{NW_{N;m} - \ln \frac{N}{m}}(t) = \frac{1}{4} \text{sech}^2\left(\frac{t}{2}\right)$$

Fig. 1 plots the scaled, asymptotic pdf (11) for various values of m . The Fermi-Dirac distribution function frequently appears in statistical physics (see e.g. [2]). It is of interest to observe that the deep tails of

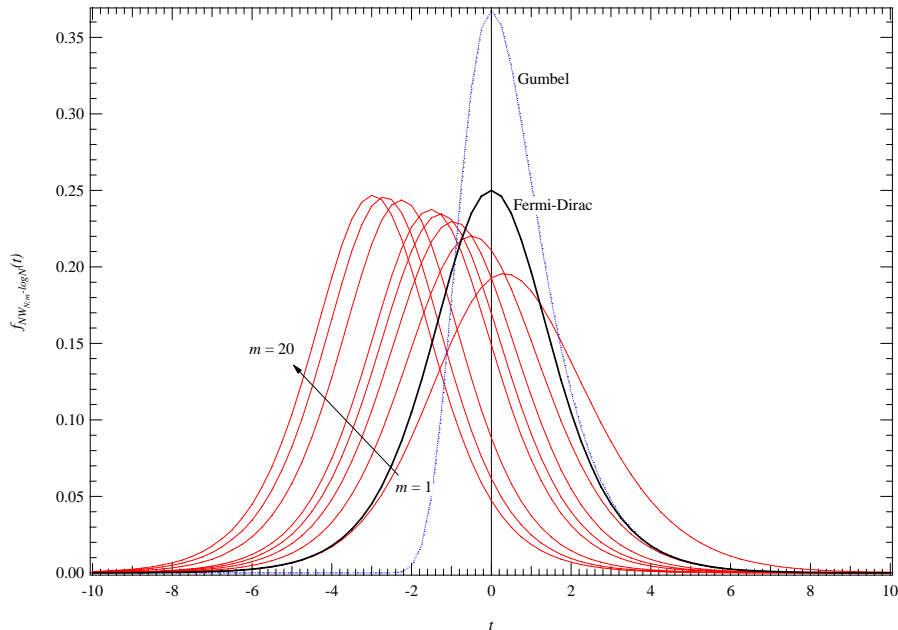


Figure 1: The asymptotic and scaled probability density function (11) for various values of $m = 1, 2, 3, 4, 5, 10, 15, 20$. The pdf of the Gumbel and the Fermi-Dirac distribution are also shown.

$f_{NW_{N;m} - \ln N}(t)$ for large $|t|$ decrease as $O(e^{-|t|})$. This is different than the decrease of either a Gumbel or Gaussian, which implies that larger variations around the mean can occur.

If we rescale (9) by changing $t = y - \ln m$, then

$$\lim_{N \rightarrow \infty} \Pr\left[NW_{N;m} - \ln \frac{N}{m} \leq y\right] = e^{-my} m^{m+1} e^{me^{-y}} \int_{me^{-y}}^{\infty} \frac{e^{-u}}{u^{m+1}} du$$

whose dependence on m and tendency to the Fermi Dirac distribution (12) is shown in Fig. 2. From a practical point of view, if $m > 5$, the much simpler Fermi-Dirac distribution is sufficiently accurate. The observation, that a peer group of $m \approx 5$ is already close to the asymptotic regime, leads us to conclude that relatively small peer group sizes suffice to offer a good service quality (e.g. small latency) and that increasing the peer group size only logarithmically (i.e. marginally) improves the quality of service of weight related metrics.

Rewriting (12) as

$$\Pr[W_{N;m} \leq y] \simeq \frac{1}{1 + \frac{N}{m} e^{-Ny}} \quad (13)$$

and after derivation, we obtain

$$f_{W_{N;m}}(y) \simeq \frac{\frac{N^2}{m} e^{-Ny}}{\left(1 + \frac{N}{m} e^{-Ny}\right)^2}$$

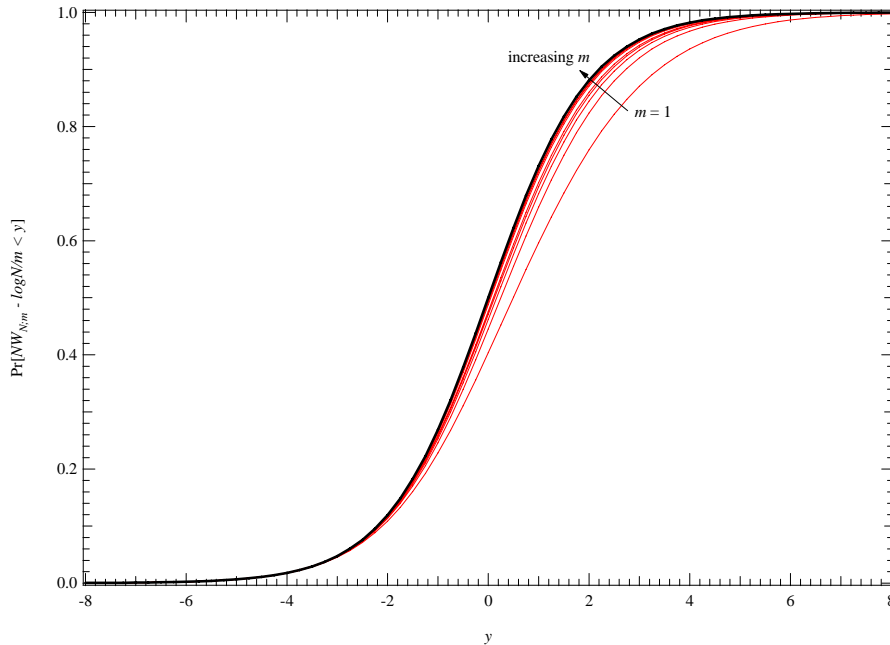


Figure 2: The tendency towards the Fermi-Dirac distribution (in bold) for increasing $m = 1, 2, 3, 4, 5, 10, 15, 20$.

whence

$$f_{W_{N;m}}(0) \simeq \frac{\frac{N^2}{m}}{\left(1 + \frac{N}{m}\right)^2} = m \left(1 + \frac{m}{N}\right)^{-2}$$

The condition that $W_{N;m} > 0$ is ignored in the asymptotic analysis, such that the largest error is made for $f_{W_{N;m}}(0)$. From (4), it follows that the accuracy of the asymptotic analysis is always better than a factor $\left(1 + \frac{m}{N}\right)^{-2}$. Hence, the smaller the fraction $\frac{m}{N}$ of peers in the network, the higher the accuracy of the asymptotic analysis. Fig. 3 shows, for finite values of N , simulation of the scaled pdf $f_{NW_{N;m} - \ln N}(t)$ and the asymptotic result (8) for $m = 1$, while Fig. 4 plots the comparison for $m = 5$. Even when ignoring the simulation errors bars, both figures indicate that the asymptotic analysis is already accurate for relatively small network size N . The accuracy factor $\left(1 + \frac{m}{N}\right)^{-2}$ also agrees and explains why the asymptotic formula (8) is less accurate for higher m .

Acknowledgement. This work has been partially supported by European Union CONTENT NoE (FP6-IST-038423) (www.ist-content.eu)

References

- [1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1968.
- [2] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt-Saunders International Editions, Tokyo, 1985.
- [3] B. C. Berndt, *Ramanujan's Notebooks*, Springer-Verlag, Part I, New York - Berlin, 1985.
- [4] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, "First-passage percolation on the random graph", *Probability Engineering Informational Science (PEIS)*, vol. 15, pp. 225-237, 2001.
- [5] P. Van Mieghem, *Performance Analysis of Communication Systems and Networks*, Cambridge University Press, 2006.
- [6] P. Van Mieghem, *Data Communications Networking*, Techne Press, Amsterdam, 2006.

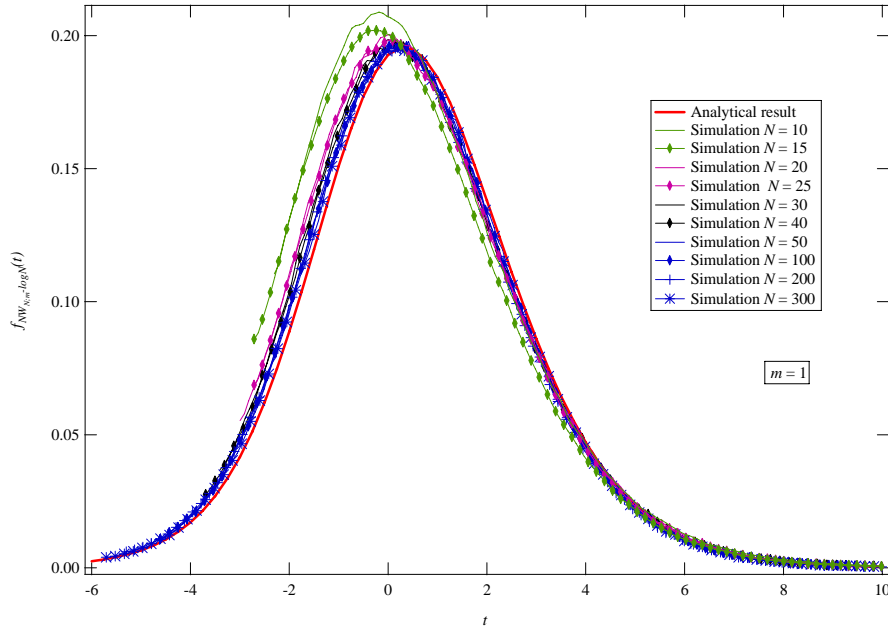


Figure 3: Comparison of the analytic result (11) with simulations for various N and $m = 1$

A Evaluation of the sum S

We introduce the Beta function [1, 6.2.1],

$$B(w, z) = \frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)} = \int_0^1 t^{z-1} (1-t)^{w-1} dt$$

which is valid for $\text{Re}(z) > 0$ and $\text{Re}(w) > 0$, in (6) such that, for $\text{Re}(x) > 0$,

$$\begin{aligned} S &= \frac{1}{\Gamma(x)} \int_0^1 dt (1-t)^{x-1} \sum_{k=1}^{N-m} t^k \frac{\Gamma(N-k+x)}{\Gamma(N-k-m+1)} \\ &= \frac{1}{\Gamma(x)} \int_0^1 dt (1-t)^{x-1} \sum_{j=m}^{N-1} t^{N-j} \frac{\Gamma(j+x)}{(j-m)!} \\ &= \frac{1}{\Gamma(x)} \int_0^1 dt (1-t)^{x-1} \sum_{j=0}^{N-m-1} t^{N-j-m} \frac{\Gamma(j+m+x)}{j!} \end{aligned}$$

Introducing Euler's Gamma integral yields

$$\begin{aligned} S &= \frac{1}{\Gamma(x)} \int_0^1 dt (1-t)^{x-1} \sum_{j=0}^{N-m-1} \frac{t^{N-m-j}}{j!} \int_0^\infty u^{j+m+x-1} e^{-u} du \\ &= \frac{1}{\Gamma(x)} \int_0^1 dt t^{N-m} (1-t)^{x-1} \int_0^\infty u^{m+x-1} e^{-u} \sum_{j=0}^{N-m-1} \frac{(u/t)^j}{j!} du \end{aligned}$$

Using (19),

$$\sum_{j=0}^{N-m-1} \frac{(u/t)^j}{j!} = \frac{e^{u/t} \Gamma(N-m, u/t)}{\Gamma(N-m)}$$

gives

$$\begin{aligned} S &= \frac{1}{\Gamma(x)\Gamma(N-m)} \int_0^1 dt t^{N-m} (1-t)^{x-1} \int_0^\infty du u^{m+x-1} e^{-u} e^{u/t} \Gamma(N-m, u/t) \\ &= \frac{1}{\Gamma(x)\Gamma(N-m)} \int_0^1 dt t^{N-m} (1-t)^{x-1} \int_0^\infty du u^{m+x-1} e^{-u} e^{u/t} \int_{u/t}^\infty y^{N-m-1} e^{-y} dy \end{aligned}$$

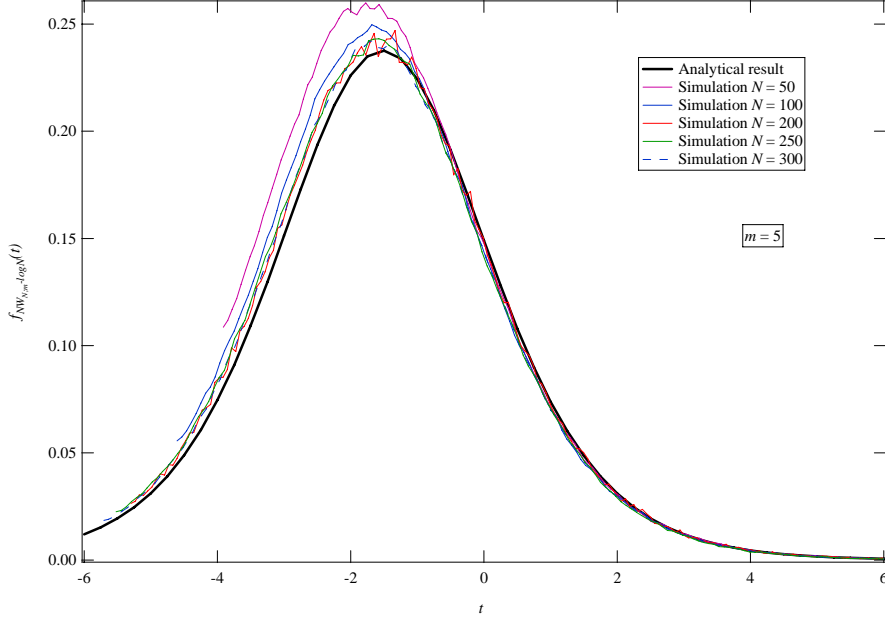


Figure 4: Comparison of the analytic result (11) with simulations for various N and $m = 5$.

Let

$$\begin{aligned} I &= \int_0^\infty du u^{m+x-1} e^{-u} e^{u/t} \int_{u/t}^\infty y^{N-m-1} e^{-y} dy \\ &= t^{m+x} \int_0^\infty dw w^{m+x-1} e^{-wt} e^w \int_w^\infty y^{N-m-1} e^{-y} dy \end{aligned}$$

Partial integration yields

$$\frac{I}{t^{m+x}} = \int_0^\infty w^{N-m-1} e^{-w} \int_0^w u^{m+x-1} e^{u(1-t)} du$$

and we find that

$$S = \frac{1}{\Gamma(x) \Gamma(N-m)} \int_0^1 dt t^{N+x} (1-t)^{x-1} \int_0^\infty w^{N-m-1} e^{-w} \int_0^w u^{m+x-1} e^{u(1-t)} du$$

We compute the u -integral as

$$\int_0^w u^{m+x-1} e^{u(1-t)} du = \sum_{k=0}^\infty \frac{(1-t)^k}{k!} \int_0^w u^{m+x-1+k} du = \sum_{k=0}^\infty \frac{(1-t)^k w^{m+x+k}}{k!(m+x+k)}$$

Substituting this series, results in

$$\begin{aligned} S &= \frac{1}{\Gamma(x) \Gamma(N-m)} \sum_{k=0}^\infty \frac{1}{k!(m+x+k)} \int_0^1 dt t^{N+x} (1-t)^{x+k-1} \int_0^\infty w^{N-1+x+k} e^{-w} dw \\ &= \frac{\Gamma(N+x+1)}{\Gamma(x) \Gamma(N-m)} \sum_{k=0}^\infty \frac{\Gamma(x+k)}{k!(m+x+k)} \frac{\Gamma(N+x+k)}{\Gamma(N+2x+k+1)} \end{aligned} \quad (14)$$

This second series (14) for S is more suited than (6) to derive the asymptotic behavior of S for large N (and fixed m and x). Invoking [1, 6.1.47] yields

$$\begin{aligned} S &= \frac{\Gamma(N+x+1) N^{-x-1}}{\Gamma(x) \Gamma(N-m)} \sum_{k=0}^\infty \frac{\Gamma(x+k)}{k!(m+x+k)} \left(1 - \frac{(x+1)(3x+2k)}{N} + O\left(\frac{k^2}{N}\right) \right) \\ &= \frac{\Gamma(N+x+1) N^{-x-1}}{\Gamma(x) \Gamma(N-m)} \left(\sum_{k=0}^\infty \frac{\Gamma(x+k)}{k!(m+x+k)} - \frac{(x+1)}{N} \sum_{k=0}^\infty \frac{(3x+2k)\Gamma(x+k)}{k!(m+x+k)} + O\left(\frac{k^2}{N^2}\right) \right) \end{aligned}$$

The remaining series can be rewritten as hypergeometric series,

$$\sum_{k=0}^{\infty} \frac{\Gamma(x+k)}{k!(m+x+k)} = \sum_{k=0}^{\infty} \frac{\Gamma(x+k)\Gamma(x+m+k)}{\Gamma(x+m+1+k)} \frac{1}{k!} = \frac{\Gamma(x)\Gamma(x+m)\Gamma(1-x)}{\Gamma(m+1)}$$

where the latter follows from Gauss's formula (16). The next sum is similarly computed with

$$\sum_{k=0}^{\infty} \frac{\Gamma(x+k)}{(k-1)!(m+x+k)} = \sum_{k=0}^{\infty} \frac{\Gamma(x+1+k)}{k!(m+1+x+k)} = \frac{\Gamma(x+1)\Gamma(x+1+m)\Gamma(-x)}{\Gamma(m+1)}$$

as

$$\frac{(x+1)}{N} \sum_{k=0}^{\infty} \frac{(3x+2k)\Gamma(x+k)}{k!(m+x+k)} = \frac{1}{N} \left(\frac{\Gamma(x+2)\Gamma(x+m)\Gamma(-x)}{\Gamma(m+1)} (2m-x) \right)$$

Higher order terms in $O(N^{-j})$ with $j > 1$ are computed similarly and converge. Combining all contributions yields

$$S = \frac{\Gamma(N+x+1)N^{-x-1}}{\Gamma(x)\Gamma(N-m)} \left(\frac{\Gamma(x)\Gamma(x+m)\Gamma(1-x)}{\Gamma(m+1)} + O(N^{-1}) \right)$$

After applying the reflection formula for the Gamma function [1, 6.1.17], $\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}$, we find (7).

B Inverse Laplace transform of (8)

It is a little easier to compute the distribution function instead of the pdf. The inverse Laplace transform is

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln N \leq t] = \frac{1}{2\pi i \Gamma(m)} \int_{c-i\infty}^{c+i\infty} \frac{\pi}{\sin \pi x} \Gamma(x+m) e^{xt} dx$$

where $0 < c < 1$. Since $\lim_{x \rightarrow -\infty} \frac{\pi x}{\sin \pi x} \frac{\Gamma(x+m)}{\Gamma(m)} e^{xt} = 0$ for any t , the contour can be closed over the negative $\text{Re}(x) < 0$ plane, where we encounter simple poles of $\frac{1}{\sin \pi x}$ at $x = -k$ and of $\Gamma(x+m)$ at $x = -m, -m-1, \dots$. Hence, the integrand has simple poles at $x = -1, -2, \dots, -m+1$ and double poles at $x = -m, -m-1, \dots$. After closing the contour over the negative $\text{Re}(x)$ -plane and applying Cauchy's residue theorem, we obtain

$$\lim_{N \rightarrow \infty} \Pr[NW_{N;m} - \ln N \leq t] = \sum_{k=0}^{m-1} \lim_{x \rightarrow -k} \frac{\pi(x+k)}{\sin \pi x} \frac{\Gamma(x+m)}{\Gamma(m)} e^{xt} + \sum_{k=m}^{\infty} \lim_{x \rightarrow -k} \left[\frac{d}{dx} \frac{\pi(x+k)^2}{\sin \pi x} \frac{\Gamma(x+m)}{\Gamma(m)} e^{xt} \right]$$

The first sum equals

$$\sum_{k=0}^{m-1} \lim_{x \rightarrow -k} \frac{\pi(x+k)}{\sin \pi x} \frac{\Gamma(x+m)}{\Gamma(m)} e^{xt} = \sum_{k=0}^{m-1} (-1)^k \frac{(m-1-k)!}{(m-1)!} e^{-kt}$$

The second sum first requires the computation of the derivative,

$$\begin{aligned} \frac{d}{dx} \frac{(x+k)^2}{\sin \pi x} \Gamma(x+m) e^{xt} &= \frac{2(x+k)}{\sin \pi x} \Gamma(x+m) e^{xt} + t \frac{(x+k)^2}{\sin \pi x} \Gamma(x+m) e^{xt} \\ &\quad - \frac{\pi(x+k)^2 \cos \pi x}{\sin^2 \pi x} \Gamma(x+m) e^{xt} + \frac{(x+k)^2}{\sin \pi x} \Gamma(x+m) \psi(x+m) e^{xt} \end{aligned}$$

Using the reflection formula of the Gamma function, $\Gamma(x+m) = \frac{\pi(-1)^m}{\sin \pi x \Gamma(1-x-m)}$, the derivative becomes

$$\frac{d}{dx} \frac{(x+k)^2}{\sin \pi x} \Gamma(x+m) e^{xt} = \frac{(x+k)^2}{\sin^2 \pi x} \frac{\pi(-1)^m}{\Gamma(1-x-m)} e^{xt} \left[\frac{2}{(x+k)} - \pi \cot \pi x + \psi(x+m) + t \right]$$

The reflection formula for the digamma function [1, 6.3.7] shows that

$$\psi(1-x-m) = \psi(x+m) + \pi \cot \pi x$$

such that the sum between brackets is

$$\frac{2}{(x+k)} - \pi \cot \pi x + \psi(x+m) + t = \frac{2}{(x+k)} - 2\pi \cot \pi x + \psi(1-x-m) + t$$

Since

$$\lim_{x \rightarrow -k} \frac{2}{(x+k)} - 2\pi \cot \pi x = 2 \lim_{x \rightarrow -k} \frac{\sin \pi x - \pi(x+k) \cos \pi x}{(x+k) \sin \pi x} = 0$$

we have

$$\lim_{x \rightarrow -k} \left[\frac{d}{dx} \frac{\pi(x+k)^2 \Gamma(x+m)}{\sin \pi x \Gamma(m)} e^{xt} \right] = \frac{(-1)^m}{(k-m)! \Gamma(m)} e^{-kt} [\psi(1+k-m) + t]$$

The second sum is

$$\begin{aligned} \sum_{k=m}^{\infty} \lim_{x \rightarrow -k} \left[\frac{d}{dx} \frac{\pi(x+k)^2 \Gamma(x+m)}{\sin \pi x \Gamma(m)} e^{xt} \right] &= \frac{(-1)^m}{\Gamma(m)} \sum_{k=m}^{\infty} \frac{e^{-kt}}{(k-m)!} [\psi(1+k-m) + t] \\ &= \frac{(-1)^m e^{-mt}}{\Gamma(m)} \sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} [\psi(1+k) + t] \\ &= \frac{(-1)^m e^{-mt}}{\Gamma(m)} \left\{ t e^{e^{-t}} + \sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} \psi(1+k) \right\} \end{aligned}$$

For integer values of the digamma function, we know [1, 6.3.2] that $\psi(1+k) = -\gamma + \sum_{j=1}^k \frac{1}{j}$, and thus,

$$\sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} \psi(1+k) = -\gamma \sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} + \sum_{k=1}^{\infty} \frac{e^{-kt}}{k!} \sum_{j=1}^k \frac{1}{j}$$

Using (21) gives

$$\sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} \psi(1+k) = -\gamma e^{e^{-t}} - e^{e^{-t}} \sum_{m=1}^{\infty} \frac{(-e^{-t})^m}{mm!}$$

The series expansion of the exponential integral $E_1(z) = \int_z^{\infty} \frac{e^{-u}}{u} du$ is [1, 5.1.11]

$$E_1(z) = -\gamma - \log z - \sum_{m=1}^{\infty} \frac{(-z)^m}{mm!}$$

and we recognize that

$$\sum_{k=0}^{\infty} \frac{e^{-kt}}{k!} \psi(1+k) = e^{e^{-t}} (E_1(e^{-t}) - t)$$

Hence,

$$\sum_{k=m}^{\infty} \lim_{x \rightarrow -k} \left[\frac{d}{dx} \frac{\pi(x+k)^2 \Gamma(x+m)}{\sin \pi x \Gamma(m)} e^{xt} \right] = \frac{(-1)^m e^{-mt}}{\Gamma(m)} e^{e^{-t}} E_1(e^{-t})$$

Combining all contributions yields

$$\lim_{N \rightarrow \infty} \Pr [NW_{N;m} - \ln N \leq t] = \sum_{k=0}^{m-1} (-1)^k \frac{(m-1-k)!}{(m-1)!} e^{-kt} + \frac{(-1)^m e^{-mt}}{(m-1)!} e^{e^{-t}} \int_{e^{-t}}^{\infty} \frac{e^{-u}}{u} du$$

Finally, after introducing (20), we arrive at (9).

C Identities

This section lists identities that we apply. For any complex a, b and positive integer k and n , there holds that

$$\sum_{j=n}^b \frac{1}{j} \frac{(a-j)!}{(b-j)!} = \frac{a!}{b!} \sum_{j=n}^b \frac{1}{j} - \frac{a!}{(b-n)!} \sum_{q=0}^{a-b-1} \frac{1}{a-b-q} \frac{(a-q-n)!}{(a-q)!} \quad (15)$$

A famous result of Gauss on the hypergeometric series is [1, 15.1.20]

$$\sum_{k=0}^{\infty} \frac{\Gamma(a+k) \Gamma(b+k)}{\Gamma(c+k) k!} = \frac{\Gamma(a) \Gamma(b) \Gamma(c-a-b)}{\Gamma(c-a) \Gamma(c-b)} \quad (16)$$

where $c \neq 0, -1, -2, \dots$ and $\text{Re}(c-a-b) > 0$.

The incomplete Gamma function is defined [1, 6.5.3] as

$$\Gamma(a, x) = \int_x^{\infty} t^{a-1} e^{-t} dt \quad (17)$$

Clearly, $\Gamma(1, x) = e^{-x}$. Partial integration yields

$$\int_x^{\infty} t^{a-1} e^{-t} dt = -\frac{x^a}{a} e^{-x} + \frac{1}{a} \int_x^{\infty} t^a e^{-t} dt$$

from which the recursion follows

$$\Gamma(a, x) = -\frac{x^a}{a} e^{-x} + \frac{1}{a} \Gamma(a+1, x) \quad (18)$$

Iterating this recursion a few times reveals that

$$\Gamma(a, x) = \frac{\Gamma(a) \Gamma(a+n, x)}{\Gamma(a+n)} - \Gamma(a) e^{-x} \sum_{k=0}^{n-1} \frac{x^{a+k}}{\Gamma(a+k+1)}$$

If $a = 1$, then

$$\sum_{k=0}^{n-1} \frac{x^k}{k!} = \frac{e^x \Gamma(n, x)}{(n-1)!} \quad (19)$$

If $a = -n$, then, after applying the reflection formula of the Gamma function, we obtain

$$\sum_{k=0}^{n-1} (-1)^k k! x^{-k-1} = e^x \left\{ \Gamma(0, x) + (-1)^{n-1} n! \Gamma(-n, x) \right\} \quad (20)$$

where $\Gamma(0, x) = E_1(x)$ is the exponential integral.

A result derived from Ramanyuan's work [3, pp. 46] is

$$\sum_{m=1}^{\infty} \frac{a_m x^m}{m} = - \sum_{m=1}^{\infty} \frac{(-1)^m f^{(m)}(x) x^m}{m!} \left(\sum_{k=1}^m \frac{1}{k} \right)$$

where $f(x) = \sum_{k=0}^{\infty} a_k x^k$ for $|x| < R$. Applied to $f(x) = e^x$ yields

$$\sum_{m=1}^{\infty} \frac{x^m}{mm!} = -e^x \sum_{m=1}^{\infty} \frac{(-x)^m}{m!} \left(\sum_{k=1}^m \frac{1}{k} \right) \quad (21)$$