C. R. Acad. Sci. Paris, Ser. I 334 (2002) 1021–1024

**Probabilités/*Probability Theory***
**(Combinatoire/*Combinatorics*)**

# The asymptotic distribution of the diameter of a random mapping *

**David Aldous, Jim Pitman**

**Department of Statistics, University of California, 367 Evans Hall # 3860, Berkeley, CA 94720-3860, USA**

**Received 17 February 2002; accepted after revision 14 March 2002**

**Note presented by Marc Yor.**

**Abstract** The asymptotic distribution of the diameter of the digraph of a uniformly distributed random mapping of an $n$-element set to itself is represented as the distribution of a functional of a reflecting Brownian bridge. This yields a formula for the Mellin transform of the asymptotic distribution, generalizing the evaluation of its mean by Flajolet and Odlyzko (1990). The methodology should be applicable to other characteristics of random mappings. ***To cite this article: D. Aldous, J. Pitman, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 1021–1024.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

**La loi limite du diamètre d'une application aléatoire**

**Résumé** On exprime la loi limite du diamètre du digraphe d'une application aléatoire, choisie uniformément parmi les applications d'un ensemble à $n$ éléments dans lui-même, comme la loi d'une fonctionnelle du pont brownien réfléchi. Ceci donne une formule pour la transformée de Mellin de cette loi limite, généralisant la formule pour sa moyenne due a Flajolet et Odlyzko (1990). Cette méthodologie devrait pouvoir s'appliquer a d'autres caractéristiques des applications aléatoires. ***Pour citer cet article : D. Aldous, J. Pitman, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 1021–1024.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

## 1. Introduction

Let $F_n$ be a uniformly distributed random mapping from the set $[n] := \{1, 2, \ldots, n\}$ to itself, as studied in [4,1] and papers cited there, and applied to pseudo-random number generators [5], and cryptography [11]. In this paper we focus on the *diameter* of $F_n$, that is the random variable $\Delta_n := \max_{i\in[n]} T_n(i)$ where $T_n(i)$ is the number of iterations of $F_n$ starting from $i$ until some value is repeated:

$$T_n(i) := \min\bigl\{ j \geqslant 1 : F_n^j(i) = F_n^k(i) \text{ for some } 0 \leqslant k < j \bigr\},$$

where $F_n^0(i) = i$ and $F_n^j(i) := F_n(F_n^{j-1}(i))$ is the image of $i$ under $j$-fold iteration of $F_n$ for $j \geqslant 1$. Flajolet and Odlyzko [4, Theorem 7] showed by singularity analysis of generating functions that

$$\lim_{n\to\infty} E\left(\frac{\Delta_n}{\sqrt{n}}\right) = \sqrt{\frac{\pi}{2}} \int_0^\infty \bigl(1 - \mathrm{e}^{-E_1(v)-I(v)}\bigr)\,\mathrm{d}v, \tag{1}$$

*E-mail address:* pitman@stat.berkeley.edu (J. Pitman).


S1631-073X(02)02386-5/FLA

where

$$E_1(v) := \int_v^\infty u^{-1}\,\mathrm{e}^{-u}\,\mathrm{d}u \quad \text{and} \quad I(v) := \int_0^v u^{-1}\,\mathrm{e}^{-u}\left[1-\exp\left(\frac{-2u}{\mathrm{e}^{v-u}-1}\right)\right]\mathrm{d}u.$$

According to our analysis [1] of the asymptotic distributions of various functionals of random mappings, there is the convergence in distribution

$$\lim_{n\to\infty} P(\Delta_n/\sqrt{n} \leqslant x) = P(\Delta \leqslant x) \tag{2}$$

for a limiting random variable $\Delta$ which can be constructed as a function of a standard Brownian bridge and a sequence of independent uniform $[0,1]$ random variables, as indicated in Section 2. The main purpose of this note is to present the following more explicit description of the law of $\Delta$, which gives probabilistic meaning to the function $\mathrm{e}^{-E_1(v)-I(v)}$ appearing in (1).

THEOREM 1.1. – *The distribution of $\Delta$ is characterized by the formula*

$$P\big(|B_1|\Delta \leqslant v\big) = \mathrm{e}^{-E_1(v)-I(v)} \quad (v \geqslant 0), \tag{3}$$

*where $B_1$ is a standard Gaussian variable independent of $\Delta$.*

COROLLARY 1.2. – *For each $p > 0$*

$$\lim_{n\to\infty} E\left[\left(\frac{\Delta_n}{\sqrt{n}}\right)^p\right] = E\big(\Delta^p\big) = \frac{p}{E(|B_1|^p)}\int_0^\infty v^{p-1}\big(1-\mathrm{e}^{-E_1(v)-I(v)}\big)\,\mathrm{d}v. \tag{4}$$

Here $E(|B_1|^p) = 2^{p/2}\Gamma((p+1)/2)/\sqrt{\pi}$, so (4) for $p=1$ reduces to (1). Formula (3) yields the second equality in (4), which characterizes the distribution of $\Delta$ by its Mellin transform. To justify the first equality in (4) we need uniform boundedness of each moment of $\Delta_n/\sqrt{n}$. But a well known bijection of Joyal bounds $\Delta_n$ by twice the height of a uniform random tree labeled by $[n]$, and the corresponding uniform boundedness for this height follows from estimates of Łuczak [6]. See also [7,9,10] for closely related Mellin transforms obtained by the technique of multiplication by a suitable independent random factor to introduce Poisson or Markovian structure. The asymptotics described here also apply to models of random mappings more general than the uniform model [2].

## 2. A Brownian bridge representation of $\Delta$

Let the connected components of the usual digraph associated with $F_n$ be put in increasing order of their least elements. For $j = 1, 2, \ldots$

- let $N_{j,n}$ be the number of elements of $[n]$ in the $j$th basin of $F_n$,
- let $C_{j,n}$ be the length of the unique cycle in the $j$th basin of $F_n$,
- let $H_{j,n}$ be the height above this cycle of the tallest tree in the $j$th basin of $F_n$.

According to [1, Theorem 8], there is convergence of finite-dimensional distributions

$$\left(\frac{N_{j,n}}{n}, \frac{C_{j,n}}{\sqrt{n}}, \frac{H_{j,n}}{\sqrt{n}}\right)_{j=1,2,\ldots} \xrightarrow{d} (\lambda_j, L_j, 2M_j)_{j=1,2,\ldots}, \tag{5}$$

where the elements in the limit can be constructed as follows from a standard Brownian bridge $B^{\mathrm{br}}$ and a sequence of independent uniform $[0,1]$ variables $U_1, U_2, \ldots$ assumed independent of $B^{\mathrm{br}}$. For $0 \leqslant v < 1$ let

$$D_v := \inf\big\{t > v : B_t^{\mathrm{br}} = 0\big\},$$

and note that $D_0 = 0$ almost surely. Let $V(0) = 0$ and let random times $V(j)$ be defined inductively as follows for $j = 1, 2, \ldots$: given that $V(i)$ has been defined for $0 \leqslant i < j$, let

$$V(j) := D_{V(j-1)} + U_j(1 - D_{V(j-1)}),$$

and let

$$\lambda_j := D_{V(j)} - D_{V(j-1)}; \quad L_j := L^{\mathrm{br}}_{D_{V(j)}} - L^{\mathrm{br}}_{D_{V(j-1)}}; \quad M_j := \max_{D_{V(j-1)} \leqslant u \leqslant D_{V(j)}} \left|B^{\mathrm{br}}_u\right|. \tag{6}$$

Since $\Delta_n = \max_j (C_{j,n} + H_{j,n})$, the asymptotic distribution of $\Delta_n/\sqrt{n}$ is the distribution of

$$\Delta := \max_j (L_j + 2M_j). \tag{7}$$

It follows easily from the construction (6), the strong Markov property of $B^{\mathrm{br}}$ at the times $D_{V(j)}$, and Brownian scaling, that

$$\lambda_j = W_j \prod_{i=1}^{j-1} (1 - W_i) \tag{8}$$

for a sequence of independent random variables $W_j$ with the beta$(1, \frac{1}{2})$ distribution $P(W_j > x) = \sqrt{1-x}$, $0 \leqslant x \leqslant 1$, and that

$$(L_j, M_j) = \sqrt{\lambda_j}\,\big(\tilde{L}_j, \widetilde{M}_j\big) \tag{9}$$

for a sequence of independent and identically distributed random pairs $(\tilde{L}_j, \widetilde{M}_j)$, independent of $(\lambda_j)$. The common distribution of $(\tilde{L}_j, \widetilde{M}_j)$ is that of

$$\big(\tilde{L}_1, \widetilde{M}_1\big) := \left(\frac{L^{\mathrm{br}}_{D_1}}{\sqrt{D_U}}, \frac{M^{\mathrm{br}}_{D_1}}{\sqrt{D_U}}\right), \tag{10}$$

where $D_U$ is the time of the first zero of $B^{\mathrm{br}}$ after a uniform[0, 1] random time $U$ which is independent of $B^{\mathrm{br}}$, and $M^{\mathrm{br}}_t := \max_{0 \leqslant u \leqslant t} |B^{\mathrm{br}}_u|$ for $0 \leqslant t \leqslant 1$. It follows from [8, Theorem 1.3] and [1, Proposition 2] that the process $B^{\mathrm{br}}_*[0, D_U]$, obtained by rescaling the path of $B^{\mathrm{br}}$ on $[0, D_U]$ to have length 1 by Brownian scaling, has the same distribution as a rearrangement of the path of the pseudo-bridge $\widetilde{B}^{\mathrm{br}} := B_*[0, \tau_1]$ where $\tau_1$ is an inverse local time at 0 for the unconditioned Brownian motion $B$. Neither the maximum nor the local time at 0 are affected by such a rearrangement, so there is the equality in distribution

$$\big(\tilde{L}_1, \widetilde{M}_1\big) \stackrel{d}{=} \big(\tilde{L}^{\mathrm{br}}, \widetilde{M}^{\mathrm{br}}\big), \tag{11}$$

where $\tilde{L}^{\mathrm{br}}$ is the local time of the pseudo-bridge $\widetilde{B}^{\mathrm{br}}$ at 0 up to time 1, and $\widetilde{M}^{\mathrm{br}} := \max_{0 \leqslant u \leqslant 1} |\widetilde{B}^{\mathrm{br}}_u|$. According to the absolute continuity relation between the laws of $\widetilde{B}^{\mathrm{br}}$ and $B^{\mathrm{br}}$ found in [3],

$$P\big(\sqrt{t}\,\tilde{L}_1 \in \mathrm{d}\ell, \sqrt{t}\,\widetilde{M}_1 \leqslant y\big) = \sqrt{\frac{2}{\pi}}\,\frac{\sqrt{t}}{\ell}\,P\big(\sqrt{t}\,L^{\mathrm{br}}_1 \in \mathrm{d}\ell, \sqrt{t}\,M^{\mathrm{br}}_1 \leqslant y\big), \tag{12}$$

for $t, \ell, y > 0$, where the joint law of $L^{\mathrm{br}}_1$ and $M^{\mathrm{br}}_1$ is characterized by the following identity [10, Theorem 3, Lemma 4 and (36)]: for all $\ell > 0$ and $y > 0$

$$\int_0^\infty \frac{\mathrm{e}^{-t/2}}{\sqrt{2\pi t}}\,\mathrm{d}t\; P\big(\sqrt{t}\,L^{\mathrm{br}}_1 \in \mathrm{d}\ell, \sqrt{t}\,M^{\mathrm{br}}_1 \leqslant y\big) = \mathrm{e}^{-\ell}\,\mathrm{d}\ell\, \exp\left(\frac{-2\ell}{\mathrm{e}^{2y} - 1}\right). \tag{13}$$

## 3. A Poisson representation of $\Delta$

It is known that for $(\lambda_j)$ as in (8), assumed independent of $B_1$, the $B_1^2\lambda_j$ are the points (in size-biased random order) of a Poisson process on $\mathbb{R}_{>0}$ with intensity measure $\frac{1}{2}t^{-1}\,\mathrm{e}^{-t/2}\,\mathrm{d}t$ which is the Lévy measure of the infinitely divisible gamma$(\frac{1}{2},\frac{1}{2})$ distribution of $B_1^2$. This yields:

LEMMA 3.1. – *If $B_1$ is a standard Gaussian variable independent of the sequence of triples $(\lambda_j, L_j, M_j)_{j=1,2,\ldots}$ featured in* (5) *and* (6)*, then the random vectors $(B_1^2\lambda_j, |B_1|L_j, |B_1|M_j)$ are the points of a Poisson point process on $\mathbb{R}^3_{>0}$ with intensity measure $\mu$ defined by*

$$\mu(\mathrm{d}t\,\mathrm{d}\ell\,\mathrm{d}m) = \frac{\mathrm{e}^{-t/2}\,\mathrm{d}t}{2t} P\big(\sqrt{t}\tilde{L}_1 \in \mathrm{d}\ell, \sqrt{t}\widetilde{M}_1 \in \mathrm{d}m\big) \tag{14}$$

*for $t,\ell,m > 0$, where $(\tilde{L}_1, \widetilde{M}_1)$ is the pair of random variables derived from a Brownian bridge by* (10)*, and the distribution of $\Delta$ defined by either* (2) *or* (7) *is characterized by the formula*

$$|B_1|\Delta = \max_j\big(|B_1|L_j + 2|B_1|M_j\big). \tag{15}$$

Using (14), (12) and (13), we deduce that the expected number of points $(|B_1|L_j, |B_1|M_j)$ with $|B_1|L_j \in \mathrm{d}\ell$ and $|B_1|M_j \leqslant y$ is

$$\int_0^\infty \frac{\mathrm{e}^{-t/2}\,\mathrm{d}t}{2t} P\big(\sqrt{t}\tilde{L}_1 \in \mathrm{d}\ell, \sqrt{t}\widetilde{M}_1 \leqslant y\big) = \ell^{-1}\,\mathrm{e}^{-\ell}\,\mathrm{d}\ell \exp\left(\frac{-2\ell}{\mathrm{e}^{2y}-1}\right). \tag{16}$$

The functions $E_1(v)$ and $I_1(v)$ featured in Theorem 1.1 can now be interpreted as follows: $E_1(v)$ is the expected number of $j$ with $|B_1|L_j \geqslant v$, while $I(v)$ is the expected number of $j$ with $|B_1|L_j < v$ and $|B_1|L_j + 2|B_1|M_j > v$. The probability of the event $|B_1|\Delta \leqslant v$, that there is no $j$ with $|B_1|L_j + 2|B_1|M_j > v$, is therefore $\mathrm{e}^{-E_1(v)-I(v)}$. The conclusion of Theorem 1.1 is now evident.

* Research supported in part by N.S.F. Grants DMS-9970901 and DMS-0071448.

## References


[1] D. Aldous, J. Pitman, Brownian bridge asymptotics for random mappings, Random Structures and Algorithms 5 (1994) 487–512.

[2] D. Aldous, J. Pitman, Invariance principles for non-uniform random mappings and trees, Technical Report 594, Dept. Statistics, U.C. Berkeley, 2001. To appear in Asymptotic Combinatorics and Mathematical Physics, Proceedings of NATO Advanced Study Institute, St Petersburg, 2001, V. Malyshev, A. Vershik (Eds.), 2002.

[3] P. Biane, J.F. Le Gall, M. Yor, Un processus qui ressemble au pont brownien, in: Séminaire de Probabilités XXI, Lecture Notes in Math., Vol. 1247, Springer, 1987, pp. 270–275.

[4] P. Flajolet, A. Odlyzko, Random mapping statistics, in: J.-J. Quisquater, J. Vandewalle (Eds.), Advances in Cryptology – EUROCRYPT '89, Lecture Notes in Comput. Sci., Vol. 434, Springer-Verlag, 1990, pp. 329–354.

[5] D.E. Knuth, The Art of Computer Programming, Vol. 2, Addison-Wesley, 1969.

[6] T. Łuckzak, Random trees and random graphs, Random Structures Algorithms 13 (1998) 485–500.

[7] M. Perman, Order statistics for jumps of normalized subordinators, Stoch. Proc. Appl. 46 (1993) 267–281.

[8] J. Pitman, M. Yor, Arcsine laws and interval partitions derived from a stable subordinator, Proc. London Math. Soc. (3) 65 (1992) 326–356.

[9] J. Pitman, M. Yor, The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator, Ann. Probab. 25 (1997) 855–900.

[10] J. Pitman, M. Yor, On the distribution of ranked heights of excursions of a Brownian bridge, Ann. Probab. 29 (2001) 362–384.

[11] D. Ye, Z. Dai, K.-Y. Lam, Decomposing attacks on asymmetric cryptography based on mapping compositions, J. Cryptology 14 (2001) 137–150.