STAPRO: 6985 Model 3Gsc pp. 1–5 (col. fig: NIL)

ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Limiting distribution of the maximal distance between random points on a circle: A moments approach

Q1 Eckhard Schlemm

University College London, United Kingdom

ARTICLE INFO

Article history:
Received 29 March 2014
Received in revised form 23 May 2014
Accepted 27 May 2014
Available online xxxx

Keywords:
Dirichlet distribution
Gumbel distribution
Maximum
Moments

ABSTRACT

Motivated by the problem of computing the distribution of the largest distance $d_{\max}$ between n random points on a circle we derive an explicit formula for the moments of the maximal component of a random vector following a Dirichlet distribution with concentration parameters $(1, \ldots, 1)$. We use this result to give a new proof of the fact that the law of $n\, d_{\max} - \log n$ converges to a Gumbel distribution as n tends to infinity.

1. Introduction

For a positive integer n, we denote by $X_1, \ldots, X_n$ a collection of n independent, standard uniform random variables, which we interpret as locations of points on a circle with perimeter one. By d_i, $i = 1, \ldots, n$, we denote the distances (in arc length) between adjacent points, that is $d_i = X_{i+1} - X_i$, where the index is taken modulo n. Alternatively, the distances d_i can be interpreted as the lengths of the pieces of a randomly broken stick of length one (Holst, 1980). A detailed understanding of their properties is of importance in some aspects of non-parametric statistics (Wilks, 1962). The set of distances is also interesting from a purely probabilistic point of view because the smallest, typical, and largest distances show quite markedly different behaviour as the number of points tends to infinity. It is known (David and Nagaraja, 2003, Problem 6.4.2) that the expected size of the kth-largest gap, $k = 1, \ldots, n$, is given by $n^{-1} \sum_{j=k}^{n} 1/j$. In particular, the smallest gap $d_{\min}$ is of order $1/n^2$, whereas the largest gap $d_{\max}$ is of order $H_n/n \sim \log n/n$, where $H_n = \sum_{j=1}^{n} 1/j$ denotes the nth harmonic number, and $a_n \sim b_n$ if and only if $a_n/b_n \to 1$. An easy calculation shows that $n^2 d_{\min}$ converges in distribution to an exponential random variable with parameter one. In this short note, we are concerned with the limiting distribution of a suitably scaled and centred version of $d_{\max}$. Using the observation that the n-tuple (d_i) of distances follows a Dirichlet distribution with parameters $(1, \ldots, 1)$ it can be deduced from Bose et al. (2008, Corollary 3.1.) that $n\, d_{\max} - \log n$ converges in law to a Gumbel distribution.

In the following we provide an alternative, combinatorical proof of that result, bypassing arguments from extreme value theory in the spirit of Gnedenko (1943) and Leadbetter et al. (1983); we derive, for the first time, an explicit formula for the moments of a $\mathrm{Dir}(1, \ldots, 1)$ distribution and compute their limits as n tends to infinity. This allows us to identify the limiting distribution in Section 3.

E-mail address: eckhard.schlemm@cantab.net.

http://dx.doi.org/10.1016/j.spl.2014.05.019
0167-7152/

Please cite this article in press as: Schlemm, E., Limiting distribution of the maximal distance between random points on a circle: A moments approach. Statistics and Probability Letters (2014), http://dx.doi.org/10.1016/j.spl.2014.05.019

2. Computation of the moments of $d_{\max}$

By David and Nagaraja (2003, Eq. (6.4.4)), the distribution function of the largest gap is given by

$$\mathbb{P}(d_{\max} \leqslant x) = 1 - n(1-x)^{n-1} + \binom{n}{2}(1-2x)^{n-1} - \cdots + (-1)^k \binom{n}{k}(1-kx)^{n-1} + \cdots,$$

where the sum continues as long as $kx \leqslant 1$. In particular, after differentiating with respect to x and observing that $d_{\max} \geqslant 1/n$, the mth moment of the largest spacing is given by

$$\mathbb{E}\left[(d_{\max})^m\right] = (n-1)\sum_{\nu=1}^{n-1}\int_{1/(\nu+1)}^{1/\nu}\sum_{k=1}^{\nu}\binom{n}{k}(-1)^{k+1}kx^m(1-kx)^{n-2}\mathrm{d}x. \tag{2.1}$$

In the following we evaluate this expression in closed form. Since the order in which the summations and integration are carried out is inconsequential, we consider the integrals in Eq. (2.1) first.

Lemma 1. *For positive integers n, $k < n$, $\nu < n$ and m, the following holds.*

$$\int_{1/(\nu+1)}^{1/\nu} x^m(1-kx)^{n-2}\mathrm{d}x = \sum_{\mu=0}^{m}\frac{1}{k^\mu}\frac{m!(n-2)!}{(m-\mu)!(n+\mu-2)!}T_{n+\mu,k,\nu,m-\mu}, \tag{2.2}$$

where

$$T_{n+\mu,k,\nu,m-\mu} = \frac{1}{k(n+\mu-1)}\left(\frac{(\nu+1-k)^{n+\mu-1}}{(\nu+1)^{n+m-1}} - \frac{(\nu-k)^{n+\mu-1}}{\nu^{n+m-1}}\right). \tag{2.3}$$

Proof. The result is obtained by m-fold integration by parts. □

After changing the order of summations, this result can be used to perform the ν-sum in Eq. (2.1).

Lemma 2. *For positive integers n, $k < n$ and m, the following holds.*

$$\sum_{\nu=k}^{n-1}\int_{1/(\nu+1)}^{1/\nu} x^m(1-kx)^{n-2}\mathrm{d}x = \sum_{\mu=0}^{m}\frac{1}{k^{\mu+1}}\frac{m!(n-2)!}{(m-\mu)!(n+\mu-1)!}\frac{(n-k)^{n+\mu-1}}{n^{n+m-1}}. \tag{2.4}$$

Proof. After plugging in Eq. (2.2) and interchanging the order of summation the sum over k is seen to be telescoping, which gives the result. □

We also need the following binomial identities whose easy proofs are left to the reader.

Lemma 3. *For positive integers n, m and $s \leqslant m$, the following identities hold.*

$$\sum_{k=1}^{n}\binom{n}{k}(-1)^{k+1}\frac{k}{n+1-k} = (-1)^{n+1}; \tag{2.5}$$

$$\sum_{k=1}^{n}k^s(-1)^{k+1}\binom{n}{k} = 0; \tag{2.6}$$

$$\sum_{\mu=s}^{m}\frac{(-1)^\mu}{(m-\mu)!(n+\mu-1)!}\binom{n+\mu-1}{\mu-s} = \delta_{s,m}\frac{(-1)^m}{(n+m-1)!}. \tag{2.7}$$

The following result records a link between raw moments and cumulants of a random variable and is used repeatedly in the sequel.

Lemma 4. *For a positive integer m and real numbers $x_1, \ldots, x_m$, the quantity*

$$\sum_{\substack{r_1+2r_2+\cdots+mr_m=m\\ r_i\in\mathbb{N}_0}}\frac{m!}{r_1!1^{r_1}\cdots r_m!m^{r_m}}x_1^{r_1}\cdots x_m^{r_m} \tag{2.8}$$

can be interpreted as $m![y^m]\exp\left\{\sum_{r=1}^m x_r y^r/r\right\}$, where $[y^m]f(y)$ denotes the coefficient of y^m in the formal power series $f(y)$. In particular, the mth cumulant of a random variable with mth raw moment given by Eq. (2.8) is equal to $\kappa_m = (m-1)!\,x_m$.