

A COMBINATORIAL IDENTITY FOR A PROBLEM IN ASYMPTOTIC STATISTICS

Hansjörg Albrecher, Jozef L. Teugels, Klaus Scheicher

Let $(X_i)_{i \geq 1}$ be a sequence of positive independent identically distributed random variables with regularly varying distribution tail of index $0 < \alpha < 1$ and define

$$T_n := \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{(X_1 + X_2 + \cdots + X_n)^2}.$$

In this note we simplify an expression for $\lim_{n \rightarrow \infty} \mathbb{E}(T_n^k)$, which was obtained by ALBRECHER and TEUGELS: *Asymptotic analysis of a measure of variation*. Theory Prob. Math. Stat., **74** (2006), 1–9, in terms of coefficients of a continued fraction expansion. The new formula establishes an unexpected link to an enumeration problem for rooted maps on orientable surfaces that was studied in ARQUÈS and BÉRAUD: *Rooted maps of orientable surfaces, Riccati's equation and continued fractions*. Discrete Mathematics, **215** (2000), 1–12.

1. INTRODUCTION

Let $(X_i)_{i \geq 1}$ be a sequence of positive independent identically distributed (i.i.d.) random variables with distribution function F . Assume that F satisfies

$$(1) \quad 1 - F(x) \sim x^{-\alpha} \ell(x) \quad \text{for } x \uparrow \infty,$$

where $\alpha > 0$ and $\ell(x)$ is slowly varying, i.e.

$$\lim_{x \rightarrow \infty} \ell(tx)/\ell(x) = 1 \quad \text{for all } t > 0$$

(cf. e.g. BINGHAM, GOLDIE and TEUGELS [4]). Relation (1) appears as the essential condition for the domain of attraction problem in extreme value theory. Note that the expectation $\mathbb{E}(X_1^\beta)$ is finite if $\beta < \alpha$ but infinite whenever $\beta > \alpha$, so

2000 Mathematics Subject Classification. Primary 05A15; Secondary 62G20.

Keywords and Phrases. Asymptotic behavior, generating functions, continued fraction, regularly varying functions, enumeration problems.

that distributions of type (1) play a crucial role for modeling extremely heavy-tailed data sets in statistics. Now define

$$(2) \quad T_n := \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{(X_1 + X_2 + \cdots + X_n)^2}.$$

This statistic turns out to have interesting properties in particular for the case $0 < \alpha < 1$ (i.e. when $\mathbb{E}(X_i)$ does not exist). Utilizing KARAMATA theory of regularly varying functions, the following asymptotic limit for arbitrary moments of T_n was shown in ALBRECHER and TEUGELS [1]:

Theorem 1. *If F satisfies (1) with $0 < \alpha < 1$, then for all $k \geq 1$*

$$(3) \quad \lim_{n \rightarrow \infty} \mathbb{E}(T_n^k) = \frac{k!}{(2k-1)!} \sum_{r=1}^k \frac{\alpha^{r-1}}{r \Gamma(1-\alpha)^r} G(r, k),$$

where $G(r, k)$ is the coefficient of x^k in the polynomial

$$(4) \quad \left(\sum_{j=1}^{k-r+1} \frac{\Gamma(2j-\alpha)}{j!} x^j \right)^r.$$

The first few moments are given by

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}(T_n) &= (1-\alpha), \\ \lim_{n \rightarrow \infty} \mathbb{E}(T_n^2) &= \frac{1}{3} (1-\alpha)(3-2\alpha), \\ \lim_{n \rightarrow \infty} \mathbb{E}(T_n^3) &= \frac{1}{15} (1-\alpha)(15-17\alpha+5\alpha^2), \\ \lim_{n \rightarrow \infty} \mathbb{E}(T_n^4) &= \frac{1}{105} (1-\alpha)(105-155\alpha+79\alpha^2-14\alpha^3) \quad \text{etc.} \end{aligned}$$

As the right-hand side is finite for each k , this result gives rise to a convenient and simple method to both estimate the extreme value index and the finiteness of the mean of a distribution in the domain of attraction of a stable law from a data set of independent and identically distributed observations. Moreover, T_n is closely connected to the study of the sample coefficient of variation and the sample dispersion (cf. ALBRECHER, LADOUCETTE and TEUGELS [2]).

Given the structure of formula (3), it is natural to ask for a simpler representation of its right-hand side through generating functions. The purpose of this note is to establish such a relationship by identifying the right-hand side as a polynomial in α with coefficients determined by a bivariate generating function of continued fraction type. Surprisingly, the result turns out to be intimately connected to the solution of an enumeration problem for rooted maps on orientable surfaces as dealt with in ARQUÈS and BÉRAUD [3].

2. AN ALTERNATIVE REPRESENTATION

Theorem 1 can be reformulated in the following way:

Theorem 2. *If F satisfies (1) with $0 < \alpha < 1$, then for all $k \geq 1$*

$$(5) \quad \lim_{n \rightarrow \infty} \mathbb{E}(T_n^k) = \frac{1}{\prod_{\ell=1}^k (2\ell - 1)} \sum_{j=0}^k (-1)^j a_{jk} \alpha^j,$$

where a_{jk} is the coefficient of $t^j z^k$ in the expansion of the continued fraction

$$(6) \quad M(t, z) = \frac{1}{1 - \frac{(t+1)z}{1 - \frac{(t+2)z}{1 - \frac{(t+3)z}{1 - \dots}}}}.$$

Proof. Define $\tau_k := \lim_{n \rightarrow \infty} \mathbb{E}(T_n^k)$. From Theorem 1 we know that

$$\frac{(2k-1)!}{k!} \tau_k = \frac{1}{\alpha} \sum_{r=1}^k \frac{1}{r} W_\alpha(r, k)$$

where $W_\alpha(r, k)$ is the coefficient of y^k in the expansion of

$$\left(\sum_{s=1}^{\infty} \frac{\alpha \Gamma(2s - \alpha)}{\Gamma(1 - \alpha)} \frac{y^s}{s!} \right)^r.$$

Turning to generating functions, we obtain

$$(7) \quad \begin{aligned} \sum_{k=1}^{\infty} \frac{(2k-1)!}{k!} \tau_k x^k &= \frac{1}{\alpha} \sum_{r=1}^{\infty} \frac{1}{r} \sum_{k=r}^{\infty} W_\alpha(r, k) x^k \\ &= \frac{1}{\alpha} \sum_{r=1}^{\infty} \frac{1}{r} \left(\sum_{s=1}^{\infty} \frac{\alpha \Gamma(2s - \alpha)}{\Gamma(1 - \alpha)} \frac{x^s}{s!} \right)^r \\ &= -\frac{1}{\alpha} \ln \left(1 - \sum_{n=1}^{\infty} \frac{\alpha \Gamma(2n - \alpha)}{\Gamma(1 - \alpha)} \frac{x^n}{n!} \right) \\ &= \frac{1}{t} \ln \left(\sum_{n=0}^{\infty} \frac{\Gamma(2n + t)}{\Gamma(t)} \frac{x^n}{n!} \right), \end{aligned}$$

where α is replaced by $-t$. Now we would like to identify the coefficients of x^k on the left-hand side of (7) as polynomials in t (α , respectively). For that purpose, we guess that

$$(8) \quad \tau_k = \frac{1}{\prod_{\ell=1}^k (2\ell - 1)} \sum_{j=0}^k a_{jk} t^j \quad \text{with} \quad a_{jk} = \frac{1}{j!k!} \mu_{jk},$$

where the terms μ_{jk} denote double partial derivatives (for $j \neq k$ twice the double partial derivatives, respectively), evaluated at zero, of some bivariate generating function $M(t, z)$, i.e.

$$(9) \quad M(t, z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} \sum_{j=0}^k \frac{t^j}{j!} \mu_{jk}$$

with $z = 2x$. If we take (8) for granted, then the generating function for the sequence $\frac{(2k-1)!}{k!} \tau_k$ can be rewritten in the form

$$(10) \quad \sum_{k=1}^{\infty} \frac{(2k-1)!}{k!} \tau_k x^k = \sum_{k=1}^{\infty} \frac{1}{2k} \frac{(2x)^k}{k!} \sum_{j=0}^k \frac{t^j}{j!} \mu_{jk},$$

From (7), (9) and (10) it follows that

$$\frac{1}{2} \int_0^x \frac{M(t, y)}{y} dy = \sum_{k=1}^{\infty} \frac{1}{2k} \frac{x^k}{k!} \sum_{j=0}^k \frac{t^j}{j!} \mu_{jk} = \frac{1}{t} \ln \left(\sum_{n=0}^{\infty} \frac{\Gamma(2n+t)}{\Gamma(t)} \frac{(x/2)^n}{n!} \right).$$

This finally leads to

$$tM(t, z) := M^*(t, z) = 2z \frac{\partial}{\partial z} \ln \left(\sum_{n=0}^{\infty} \frac{\Gamma(2n+y)}{\Gamma(y)} \frac{(z/2)^n}{n!} \right).$$

But, by algebraic techniques, $M^*(t, z)$ was identified in JACKSON and VISENTIN [5, Prop. 3.6] as the generating function for all rooted maps on orientable surfaces, without regard to genus, with respect to edges and vertices. In [3], using a topological approach, ARQUÈS and BERAUD alternatively identified this generating function as the solution of the RICCATI differential equation

$$(1 - z(2t + 1)) M^*(t, z) = z M^*(t, z)^2 + z(t^2 + t) + 2z^2 \frac{\partial}{\partial z} M^*(t, z),$$

for which they gave the solution in terms of the continued fraction

$$M^*(t, z) = \frac{t}{1 - \frac{(t+1)z}{1 - \frac{(t+2)z}{1 - \frac{(t+3)z}{1 - \dots}}}} - t.$$

The summand $-t$ above can be omitted, since we are only interested in terms of the expansion for which the power of z is at least one (corresponding to $k \geq 1$). Hence we finally arrive at the desired result. \square

REMARK. The expression $G(r, k)$ of Theorem 1 above is a consequence of its original form

$$G(r, k) = \sum_{\substack{k_1, \dots, k_r \geq 1 \\ k_1 + \dots + k_r = k}} \prod_{j=1}^r \frac{\Gamma(2k_j - \alpha)}{k_j!},$$

which appeared by collecting all asymptotically relevant terms of the multinomial expansion of an integral representation of $\mathbb{E}(T_n^k)$ (see [1] for details). It is somewhat surprising that the resulting counting problem in (4) is intimately connected with the problem of counting all possible orientable rooted maps of any genus for a given number of edges and vertices.

REFERENCES

1. H. ALBRECHER, J. TEUGELS: *Asymptotic analysis of a measure of variation*. Theory Prob. Math. Stat., **74** (2006), 1–9.
2. H. ALBRECHER, S. LADOUCKETTE, J. TEUGELS: *Asymptotics of the sample coefficient of variation and the sample dispersion*. K. U. Leuven, Preprint (2008).
3. D. ARQUÈS, J.-F. BÉRAUD: *Rooted maps of orientable surfaces, Riccati's equation and continued fractions*. Discrete Mathematics, **215** (2000), 1–12.
4. N. BINGHAM, C. GOLDIE, J. TEUGELS: *Regular variation*. Cambridge University Press, Cambridge, (1987).
5. D. M. JACKSON, T. I. VISENTIN: *A character theoretic approach to embeddings of rooted maps in an orientable surface of given genus*. Transactions of the American Mathematical Society, **322** (1990), 343–363.

HANSJÖRG ALBRECHER¹

Institute of Actuarial Science,
Faculty HEC, University of Lausanne,
Extranef Building, CH-1015 Lausanne,
Switzerland

E-mail: hansjoerg.albrecher@unil.ch

(Received October 9, 2008)

(Revised December 12, 2008)

KLAUS SCHEICHER²

University of Natural Resources and Applied Life Sciences,
Gregor Mendel Strasse 33, 1180 Vienna, Austria

E-mail: klaus.scheicher@boku.ac.at

JOZEF L. TEUGELS

EURANDOM, Technische Universiteit Eindhoven,
The Netherlands and Katholieke Universiteit Leuven,
Leuven Center for Statistics,
Celestijnenlaan 200B, B-3001 Heverlee, Belgium

E-mail: jef.teugels@wis.kuleuven.be

¹Supported by the Austrian Science Fund Project P18392

²Supported by the FWF research projects P18079-N12 and P20989