The Ratio of the Extreme to the Sum in a Random Sequence with Applications

Peter J. Downey

The University of Arizona

Paul E. Wright

AT&T Bell Laboratories

TR 94-18

ABSTRACT

If $X_1, X_2, \cdots, X_n$ is a sequence of non-negative independent random variables with common distribution function $F(t)$, we write $X_{(n)}$ for the maximum of the sequence and S_n for its sum. The ratio variate $R_n = X_{(n)}/S_n$ is a quantity arising in the analysis of process speedup and the performance of scheduling. O'Brien (1980) showed that $R_n \to 0$ almost surely $\Leftrightarrow$ $\mathbf{E}X_1 < \infty$ as $n \to \infty$. Since $\{R_n\}$ is a uniformly bounded sequence it follows that $\mathbf{E}X_1 < \infty \Rightarrow \mathbf{E}R_n \to 0$ as $n \to \infty$.

Here we show that, provided either that (i) $\mathbf{E}X_1^2 < \infty$ or that (ii) $1 - F(t)$ is a regularly varying function with index $\rho < -1$, it follows that

$$\mathbf{E}R_n = \frac{\mathbf{E}X_{(n)}}{\mathbf{E}S_n}\left[1 + o(1)\right] \qquad (n \to \infty) .$$

Since the asymptotics of $\mathbf{E}X_{(n)}$ is often readily calculated, this provides a useful estimate for the most significant behavior of the ratio R_n in expectation. We apply this result to multiprocessor scheduling policies and to the behavior of sample statistics.

June 28, 1994

Department of Computer Science

The University of Arizona

Tucson, Arizona 85721

The Ratio of the Extreme to the Sum in a Random Sequence with Applications

1. Introduction and Results

Let $X_1, X_2, \cdots, X_n$ be a sequence of independent non-negative random variables (r.v.) with the common distribution function (d.f.) $F(t)$. The *extreme* and the *sum* of the sequence are the r.v.s

$$X_{(n)} := \max(X_1, X_2, \cdots, X_n) , \qquad S_n := \sum_{i=1}^{n} X_i .$$

How influential a contribution does the extreme term $X_{(n)}$ make to the sum S_n for large n? Its relative contribution is the ratio

$$R_n := \frac{X_{(n)}}{S_n} .$$

It is not difficult to see that if F has finite mean, R_n becomes asymptotically negligible: $R_n \to 0$ almost surely and $\mathbf{E}R_n \to 0$ as well. O'Brien [OBr80] gave necessary and sufficient conditions on F for R_n to converge to zero almost surely and in probability, and that characterize convergence in moment. It will be seen from applications below that is it is useful to be able to quantify the rate of approach to zero. One way to do this is to express $\mathbf{E}R_n$ in terms of the simpler expectations $\mathbf{E}X_{(n)}$ and $\mathbf{E}S_n$. Of course, the expected ratio is in general not the ratio of the latter expectations—because R_n and S_n are typically correlated random variables. Nevertheless, we will show in this paper that for many distributions F this relationship is *asymptotically* true. In Theorem (6.1) we provide mild sufficient conditions on F that imply

$$\mathbf{E}R_n \sim \frac{\mathbf{E}X_{(n)}}{\mathbf{E}S_n} = \frac{\mathbf{E}X_{(n)}}{\mu n} \qquad (n \to \infty) ,$$

where $\mu := \mathbf{E}X_1$. Whenever the theorem applies, it serves to reduce estimation of $\mathbf{E}R_n$ to the easier task of estimating $\mathbf{E}X_{(n)}$, a quantity that has been well studied.

After some preliminary definitions, we review the known results on the asymptotic behavior of R_n. We then state our main result, and provide some illustrative applications. The bulk of the paper is then devoted to the proof of Theorem (6.1).

1.1 Definitions

Unless otherwise indicated, all random variables are non-negative. If random variables X and Y are identically distributed, we write $X =_d Y$. Similarly, we use the shorthand $X =_d F$ to mean that X has distribution function F. The complementary distribution function $1 - F(x)$ is denoted $\overline{F}(x)$. To avoid trivialities, we will assume that F is not a single atom.

Expectations are denoted by $\mathbf{E}$. For non-negative variates, $\mathbf{E}X$ exists and is either finite or infinite. For a non-negative random variable $X =_d F$

$$\mathbf{E}X^p = \int_0^\infty x^p \, dF(x) = p \int_0^\infty x^{p-1} \overline{F}(x) dx . \tag{1.1.1}$$

where the integrals are both finite or both infinite together. If $\mathbf{E}[X^p]$ is finite, so is $\mathbf{E}[X^q]$ for all $q < p$.

Define the truncated moments of $X =_d F$ by

$$\mu_i(z) := \int_0^z t^i \, dF(t)$$

We write $\mu(z)$ for the truncated mean $\mu_1(z)$, and often write μ (= $\mu(\infty)$) for the mean.

Let f and g be functions. The relation of approximate dominance, denoted $f(x) = O(g(x))$, means that the ratio $|f(x)/g(x)|$ is bounded as $x \to \infty$. The strong version of this relation, denoted $f(x) = o(g(x))$, means that the ratio $|f(x)/g(x)|$ goes to zero as $x \to \infty$. The relation of asymptotic equality, denoted $f(x) \sim g(x), \quad (x \to \infty)$, means that the ratio $f(x)/g(x)$ tends to 1 in the limit of large x. The relation of approximate asymptotic equality, denoted $f(x) = \Theta(g(x))$ means that, as $x \to \infty$, the ratio $f(x)/g(x)$ is contained in some closed positive interval $[l, u]$, $0 < l \le u < \infty$.

If $X_1, X_2, \cdots$ is a sequence of r.v.s, and X is an r.v., then the notations

$$X_n \xrightarrow{as} X, \quad X_n \xrightarrow{pr} X, \quad X_n \xrightarrow{d} X, \quad \text{and} \quad \mathbf{E}X_n \to \mathbf{E}X$$

denote that the sequence converges *almost surely*, *in probability*, *in distribution* and *in moment*, respectively. For definitions see [Chu74].

If f is a positive function, it said to be *regularly varying with index* ρ ($f \in R_\rho$) iff for all constants $c > 0$

$$f(ct) \sim c^\rho f(t) \qquad (t \to \infty) .$$

Functions in R_0 are called *slowly varying*. A fuller account of regular variation appears in section 2.1 below.

1.2 History

We begin with results characterizing when R_n vanishes for large n. As is the case throughout, we assume that $X_1, \cdots, X_n$ are independent and identically distributed (i.i.d.) with d.f. F.

THEOREM [OBr80]. Assume $X_1, \cdots, X_n$ are i.i.d. with d.f. F. The following are equivalent:

(i) $R_n \xrightarrow{as} 0$ (1.1.2)

(ii) $\mathbf{E}X_1 < \infty$. □

A slightly larger class of distributions allows for weak convergence:

THEOREM [OBr80]. Assume $X_1, \cdots, X_n$ are i.i.d. with d.f. F. The following are equivalent:

(i) $R_n \xrightarrow{pr} 0$ (1.1.3)

(ii) $\mu(x) \in R_0$, i.e., the truncated mean varies slowly as $x \to \infty$

(iii) $\mathbf{E}R_n \to 0$.

Proof: (i) $\Leftrightarrow$ (ii) is proved in [OBr80]. Convergence in mean always implies convergence in probability [Chu74], so (iii) $\Rightarrow$ (i). Conversely, since the sequence R_n is uniformly bounded by 1 and converges in probability to zero, then it converges in mean as well [Chu74]. This shows (i) $\Rightarrow$ (iii). □

Another necessary and sufficient condition for (i) is given in [Bre65].

EXAMPLE. The d.f. given by $F(t) = 1 - t^{-1}$ on $[1, \infty)$ has an infinite mean; however $\mu(x) = \ln x - 1$ is slowly varying and so Theorem (1.1.3) applies, yet the sequence R_n has a limit superior of 1 almost surely. □

If a d.f. has no finite moments of any order $\varepsilon > 0$, the extreme term $X_{(n)}$ is of the same order as the entire sum. This counterintuitive state of affairs was characterized in [Mal84], building on prior work of [Dar52].

THEOREM [Mal84]. Assume $X_1, \cdots, X_n$ are i.i.d. with d.f. F. The following are equivalent:

(i) $R_n \xrightarrow{pr} 1$ (1.1.4)

(ii) $\bar{F}(t) \in R_0$. □

Note that (i) implies $\mathbf{E}R_n \to 1$. A somewhat technical necessary and sufficient condition for $R_n \to 1$ almost surely is given in [Pru87].

EXAMPLE. The d.f. $F(t) = 1 - (\ln t)^{-1}$ defined on $[e, \infty)$ is such that $\bar{F}(t)$ is slowly varying. One sees that $\mathbf{E}[X^\varepsilon] = \infty$ for all $\varepsilon > 0$. Here R_n converges to 1 in probability and expectation for large n. □

For d.f.s F that fail to have a finite mean, but have finite moments of some fractional order, the maximum term $X_{(n)}$ can make up a fixed proportion of the sum. Prior work of [Dar52, Aro60, Bre65] culminates in

THEOREM [Bin81]. Assume $X_1, \cdots, X_n$ are i.i.d. with d.f. F. The following are equivalent:

(i) $R_n \xrightarrow{d} R$ where R has a non-degenerate distribution; (1.1.5)

(ii) $\bar{F}(t) \in R_{-\alpha}$ for some $\alpha \in (0,1)$,

(i.e., F is attracted to a stable law of index α in $(0,1)$)

(iii) $\mathbf{E}[S_n/X_{(n)}] \to \dfrac{1}{1-\alpha}$ for some $\alpha \in (0,1)$. □

EXAMPLE. In the case $\alpha = ½$, [Bre65] shows that the sequence of distributions of the variates R_n converges to the limiting d.f. of R given by the $\beta_{½,½}$ distribution $\mathbf{P}[R \le t] = (2/\pi)\arcsin(\sqrt{t})$. □

An analogous result to (1.1.5) provides information about $\mathbf{E}[1/R_n]$ for regularly varying d.f.s having finite mean. Extending earlier results of [Aro60, Dar52, Cho79], Bingham and Teugels showed:

THEOREM [Bin81]. Assume $X_1, \cdots, X_n$ are i.i.d. with d.f. F. The following are equivalent:

(i) $[S_n - (n-1)\mu]/X_{(n)} \xrightarrow{d} D$ where D has a non-degenerate distribution; (1.1.6)

(ii) $\bar{F}(t) \in R_{-\alpha}$ for some $\alpha \in (1,2)$,

(i.e., F is attracted to a stable law of index α in $(1,2)$)

(iii) $\mathbf{E}[[S_n - (n-1)\mu]/X_{(n)}] \to c$ for some constant c . □

Unlike (1.1.5), where $X_{(n)}$ accounts for a fixed proportion of the sum S_n, (1.1.6) gives conditions under which $X_{(n)}$ accounts for a declining portion of S_n as n increases. Indeed, we give a complementary view of this in the main Theorem (6.1) of this paper, which implies that for regularly varying d.f.s in this range of index $\mathbf{E}R_n \sim \mathbf{E}X_{(n)}/(n\mu)$.

1.3 Main Theorem

For certain F with finite first moment, we can quantify the approach to zero in expectation via the following result. The proof is deferred to section 6.

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$. Suppose either that

(i) $\bar{F} \in R_{-\alpha}$ for some $\alpha > 1$; or else

(ii) F has finite second moment.

Then

$$\mathbf{E}R_n = \frac{\mathbf{E}X_{(n)}}{\mathbf{E}S_n}\Big[1 + o(1)\Big] \qquad (n \to \infty) . \qquad \square \tag{6.1}$$

REMARK. The hypotheses of Theorem (6.1) are familiar in classical probability theory: they are the necessary and sufficient conditions for d.f. F to be in the domain of attraction of a stable distribution [Fel66, XVII.5, IX.8]. We briefly review these domain of attraction conditions here.

(I) F is attracted to a normal limit law if and only if either

(a) $\mathbf{E}[X_1^2] < \infty$ (the usual Central Limit Theorem); or

(b) $\bar{F} \in R_{-\alpha}$ for $\alpha \in [2, \infty]$; or

(c) the truncated second moment $\mu_2(x)$ is slowly varying.

In fact, condition (c) subsumes both (a) and (b) [Fel66, VIII.9], so that F is attracted to a normal stable law $\Leftrightarrow \mu_2 \in R_0$.

(II) F is attracted to a non-normal stable law of index $\alpha \in (0,2) \Leftrightarrow \bar{F} \in R_{-\alpha}$ for $\alpha \in (0,2)$.

In spite of this intimate connection with stability, the proofs in this paper make no use of stable attraction explicitly, nor do we know that the hypotheses in (6.1) are necessary for its conclusion. This remains a tantalizing open question.

1.4 Applications

Theorem (6.1) has application in contexts where the ratio variate R_n naturally arises. Some illustrations are given below. In this section, we will assume that F obeys the hypotheses of Theorem (6.1): either (i) $\bar{F} \in R_{-\alpha}$ for some $\alpha > 1$ or else (ii) F has finite second moment.

1.4.1 Sample Statistics

Among measures of dispersion of a distribution F, the most common is the coefficient of variation (c.v., relative standard deviation) σ/μ. Consider a sample $X_1, \cdots, X_n$ of i.i.d. r.v.s from F, and define the *sample c.v.* as the random variable

$$\frac{s_n}{\bar{X}_n} := \frac{\sqrt{(n-1)^{-1}\sum_i (X_i - \bar{X}_n)^2}}{\bar{X}_n}$$

where $\bar{X}_n := S_n/n$ is the sample mean. Then, provided σ is finite, the strong law of large numbers implies that the sample c.v. converges a.s. to the population c.v. σ/μ. It also conveges in moment, so that the sample c.v. is unbiased.

Other sample measure of relative dispersion have been used that are based upon the extremes of the sample. They include the relative deviation [Dav70]

$$RD_n := \frac{X_{(n)} - \bar{X}_n}{\bar{X}_n} ,$$

the relative center [Dav70]

$$RC_n := \frac{X_{(n)} + X_{(1)}}{2\bar{X}_n} ,$$

where $X_{(1)}$ is the minimum sample value, the relative range [Dav70]

$$RR_n := \frac{X_{(n)} - X_{(1)}}{\bar{X}_n} ,$$

and the peak-to-mean ratio

$$PM_n := \frac{X_{(n)}}{\bar{X}_n} .$$

Now rather than being estimates of dispersion, all these statistics are asymptotically estimates of $X_{(n)}$:

THEOREM. Under hypothesis (i) or (ii), if F has unbounded support, then

$$\mathbf{E}RD_n \sim 2 \cdot \mathbf{E}RC_n \sim \mathbf{E}RR_n \sim \mathbf{E}PM_n \sim \frac{\mathbf{E}X_{(n)}}{\mu} \qquad (n \to \infty) . \tag{1.4.1.1}$$

Proof: The proof for RC_n illustrates all the others. $2 \cdot RC_2 = nR_n + nX_{(1)}/S_n$. Use the observation $nX_{(1)} \le S_n$ and Theorem (6.1) to get $2 \cdot \mathbf{E}RC_n \sim \mathbf{E}X_{(n)}/\mu$ since $\mathbf{E}X_{(n)}$ goes to infinity. □

Any information contained in these statistics is equivalent to using $X_{(n)}$ for large samples, and it is well known that $X_{(n)}$ is sensitive only to the behavior of the upper tail of F.

The result concerning $\mathbf{E}PM_n$ quantifies the influence on $\bar{X}_n$ of the outlier term $n^{-1}X_{(n)}$. The relative influence is $\mathbf{E}PM_n/n \sim \mathbf{E}X_{(n)}/(n\mu)$. This is a decreasing function of n; see Table 1 below for estimates on its rate of decay. For distributions with finite variance, this influence decays like $o(n^{-1/2})$; and for distributions with upper tails that decay exponentially, the decay is $O(\ln n/n)$. This allows quantification of the effect of eliminaton of sample outliers upon estimates of $\bar{X}_n$.

1.4.2 Multiprocessor Makespan Scheduling

Let $\bar{X} = (X_1, \cdots, X_{n)}$ be a sequence of non-negative real numbers denoting the service times of n tasks. Suppose there are $m \geq 2$ processors that serve these tasks in parallel. Service of a task is non-preemptive: a task that is started must be run to completion. A *schedule* [Cof76] is an assignment of tasks to processors. A schedule is *conservative* if no processor is allowed to be idle while some task remains unassigned. One large class of conservative schedules are the *list* schedules, defined as follows. A *list* is a permutation L of $\{1, \cdots, n\}$. Given a list L, the list schedule (also denoted L) assigns tasks to processors in the list order $X_{L(1)}, \cdots, X_{L(n)}$. With $n > m$, the m processors begin by serving $X_{L(1)},..., X_{L(m)}$. Whenever a processor completes a task, a new task is assigned to it from the head of the remaining list $X_{L(m+1)}, \cdots, X_{L(n)}$ of unassigned tasks. After $n+1-m$ task completions, some processors must be idle. The *makespan* of schedule L is the time from the beginning of service until all tasks have completed. The makespan is a function of the task times, the schedule L, and the number of processors m, and will be denoted $M(L, m, \bar{X})$.

Ideally we would like to construct a list schedule that will minimize $M(L, m, \bar{X})$ over all possible permutations L. This optimal list schedule is denoted OPT. Calculating OPT is the *makespan scheduling problem* [Gra76]. Given explicit values for $\bar{X}$ and m, it is an NP-hard problem to compute OPT [Gar79]. The difficulty of computing the exact optimum OPT motivates the analysis of simple suboptimal list scheduling rules: how far away from optimum are they?

The *performance ratio* $M(L, m, \bar{X})/M(OPT, m, \bar{X})$ measures the penalty paid for using list schedule L. Graham [Gra69] showed that for any list schedule L and any $\bar{X}$

$$\frac{M(L, m, \bar{X})}{M(OPT, m, \bar{X})} \leq 2 - \frac{1}{m} .$$

Let LPT be the *largest processing time first* permuation that orders the tasks in decreasing order of their times: $X_{L(1)} \geq X_{L(2)} \geq \cdots \geq X_{L(n)}$. Then Graham [Gra69] also showed

$$\frac{M(LPT, m, \bar{X})}{M(OPT, m, \bar{X})} \leq \frac{4}{3} - \frac{1}{3m} .$$

The above results and others [Law93] are combinatorial and deterministic, assuming full knowledge of the set of task times $X_1, \cdots, X_n$. More recent work [Cof91] has focussed upon the following stochastic problem. Assume the task times X_i are independent and identically distributed with distribution F. How does the performance ratio variate $M(L, m, \bar{X})/M(OPT, m, \bar{X})$ behave, especially for large n? Coffman and Gilbert [Cof85] gave a bound, valid for any list schedule L, for the expected performance ratio in the case when F is a uniform or exponential law. For both these distributions, $\mathbf{E}[M(L, m, \bar{X})/M(OPT, m, \bar{X})] = 1 + O(1/n)$ for large n and any list schedule L. Other stochastic bounds for the LPT schedule are reviewed in [Cof91].

Using Theorem (6.1), we can give asymptotic bounds on the expected performance ratio for a large class of distribution functions for task times.

Before proceeding to derive these general bounds, we present a useful combinatorial identity that bounds the makespan.

THEOREM [Gra76, Theorem 5.3]. Let n tasks have processing times $\bar{X} = (X_1, \cdots, X_n)$, and let $X_{(n)}$ denote their maximum and S_n their sum. Then for any *list* schedule L assigning the tasks to m processors,

$$\max\left\{ \frac{S_n}{m}, X_{(n)} \right\} \leq M(L, m, \bar{X}) \leq \frac{S_n}{m} + \frac{m-1}{m} X_{(n)} . \quad \square \tag{1.4.2.1}$$

Now OPT is a list schedule, so that the lower bound above applies to it, and $M(OPT, m, \bar{X}) \geq S_n/m$. Dividing this into (1.4.2.1) yields:

$$1 \leq \frac{M(L, m, \bar{X})}{M(OPT, m, \bar{X})} \leq 1 + (m-1)R_n . \tag{1.4.2.2}$$

Suppose now that the task times $\bar{X}$ are chosen i.i.d $=_d F$ with mean μ. The quantities M, OPT, and R_n become random variables. From (1.4.2.2) and O'Brien's result that $R_n \to 0$ almost surely, we have that

$$\frac{M(L,\, m,\, \bar{X})}{M(OPT,\, m,\, \bar{X})} \;=\; 1 + o(1) \;\; \text{a.s.} \qquad (n \to \infty) \;. \tag{1.4.2.3}$$

Taking expectations in (1.4.2.2) and using Theorem (6.1) allows us to quantify the expected rate of convergence to unity as:

$$1 \;\le\; \mathbf{E}\left[\frac{M(L,\, m,\, \bar{X})}{M(OPT,\, m,\, \bar{X})}\right] \;\le\; 1 + (m-1)\frac{\mathbf{E}X_{(n)}}{\mu n}\left[1 + o(1)\right] \qquad (n \to \infty) \;. \tag{1.4.2.4}$$

It is known [Dow94] that $\mathbf{E}X_{(n)}/n$ is a completely monotone function (decreasing to zero and convex) of n. Since $\mathbf{E}X_{(n)}/n = o(1)$ from Theorem (2.2.1) below, this result complements (1.4.2.3), and allows us to state that under hypotheses (i) or (ii) the relative performance of any list schedule approaches optimal, almost surely and in expectation, as n gets large.

More can be said, however. Better bounds on the growth of $\mathbf{E}X_{(n)}$ with n allow us to use (1.4.2.4) to quantify the approach of the performance ratio to 1. If hypothesis (i) holds and $X \in R_{-\alpha}$, then $\mathbf{E}X_{(n)}/n \sim \Gamma(1-\alpha^{-1}) \cdot n^{1/\alpha - 1}$. If hypothesis (ii) holds and $\mathbf{E}X^2$ is finite, then it is $o(n^{-1/2})$. These and other bounds on $\mathbf{E}X_{(n)}/n$ are summarized in Table 1 below.

2. Preliminaries

In this section are collected most of the supporting lemmas used in the main arguments beginning in Section 3. Material in this section can be referred back to as needed in the sequel.

2.1 Regular Variation

Regularly varying functions are those that scale homogeneously for large argument. In this paper we are interested in distribution functions $F(t)$ for which the complementary d.f. $1 - F(t)$ is a regularly varying function with negative index. However, we will begin with the general definition of a regularly varying function, and cite the general properties of such functions that are needed later.

DEFINITION. A positive measurable function $f : (0, \infty) \to (0, \infty)$ is *regularly varying* at infinity if for all $\lambda > 1$, the limit

$$\lim_{x \to \infty} \frac{f(\lambda x)}{f(x)} \tag{2.1.1}$$

exists and is in $(0, \infty)$. f is *rapidly varying* if the limit exists and is 0 or ∞.

The fundamental result about regular variation [Bin87, Theorem 1.4.1] is that *if* the (finite or infinite) limit (2.1.1) exists for all $\lambda > 1$, then there is an extended real number ρ, $-\infty \le \rho \le \infty$ such that

$$\forall \lambda > 0 \qquad \lim_{x \to \infty} \frac{f(\lambda x)}{f(x)} = \lambda^{\rho} \;. \tag{2.1.2}$$

This ρ is called the *exponent* or *index of variation*. If (2.1.2) holds, so that f is regularly varying with index ρ, we write $f \in R_{\rho}$. Thus with the understanding $\lambda^{-\infty} = 0$ and $\lambda^{\infty} = \infty$, $R_{-\infty}$ and R_{∞} are the rapidly varying functions.

Functions like $x^a + \sin x$ where $a > 0$ are in R_a. Functions like $\exp(-x^k)$, $k > 0$ belong to $R_{-\infty}$ and their reciprocals belong to R_{∞}. Functions that fail to have a well-defined limit (finite or infinite) in (2.1.2) are neither regularly varying nor rapidly varying. Examples are $\sin x$ and $e^{-\lfloor x \rfloor}$. By historical convention, the class R_0 is called the *slowly varying* functions. It includes, for example functions like $(\ln x)^{\alpha}$ for any $\alpha \ge 0$, and their reciprocals.

From the results above, it is evident that $f \in R_{\rho}$ if and only if there is some slowly varying function l such that $f(x) = x^{\rho} \cdot l(x)$. It is easy to see that if $f \in R_{\rho}$ and $g \sim f$ then $g \in R_{\rho}$. Thus asymptotic equality $\sim$ is the natural equivalence relation on the class of regularly varying functions.

In applications to random variables, we say that $X =_d F$ is a *regularly varying random variable of index* $-\alpha$, for some $\alpha \ge 0$, provided $\bar{F}(t) \in R_{-\alpha}$. In this case we write $X \in R_{-\alpha}$. For example, if $X =_d F(t)$ where $F(t) = 1 - (\ln t)^{-1}$ for $t \in [e, \infty)$, then $X \in R_0$.

d.f. hypotheses	$\mathbf{E}X_{(n)}$ *bound*	$\mathbf{E}X_{(n)}/n$ *bound*	*Note*
$\mathbf{E}X < \infty$	$o(n)$	$o(1)$	1
$\mathbf{E}X^p < \infty\ (p > 1)$	$o(n^{1/p}),\ \mu + \lVert X - \mu \rVert_p \cdot n^{1/p}$	$o(n^{1/p-1})$	2
$X \in R_{-\alpha}\ (\alpha > 1)$	$\sim \Gamma(1-\alpha^{-1}) c_X(n)\ \ (c_X(n) \in R_{1/\alpha})$	$\sim \Gamma(1-\alpha^{-1}) c_X(n)/n$	3
$X \in R_{-\infty}$	$\sim c_X(n)\ \ (c_X(n) \in R_0)$	$\sim c_X(n)/n$	4
$X \in E^1$	$O(\ln n)$	$O(\ln n/n)$	5
$X \in E^\beta$	$O(\ln n)^\beta$	$O((\ln n)^\beta/n)$	6

Asymptotic Bounds for $n \to \infty$

Notes:

(1). See Theorem 2.2.1 below. An example d.f. in this class is $\overline{F}(t) = t^{-1}(\ln t)^{-2}$ for $t \geq \varepsilon$.

(2). See [Arn85] and [Dow90a]. The o bound assumes a fixed parent d.f. F that is independent of n, while the (weaker) bound allows F to depend on n, but has the advantage of explicit coefficients. Here $\lVert Y \rVert_p := (\mathbf{E}[Y^p])^{1/p}$ is the L^p norm.

(3). The characteristic maximum function $c_X(n)$ is (roughly) the solution to $\overline{F}(x) = n^{-1}$; see below at (2.2.7). An example is the empirical distribution of Unix process times [Lel86], found to be $\overline{F}(t) = 0.241 t^{-1.122}$. This yields $c_X(n) = 0.281 n^{0.891}$ and so $\mathbf{E}X_{(n)}/n \sim 2.452 n^{-0.109}$. For the Unix workload, this is a rather slowly decaying bound, owing to the fat tail of this d.f. As another example, if $\overline{F}(t) = t^{-2}(\ln t)^{-1}$ for $t \geq e$, then $\mathbf{E}X_{(n)} \sim \sqrt{\pi n / \ln n}$.

(4). Genedenko [Gne43] showed $X_{(n)} \sim c_X(n)$ almost surely in this case; for the expectation see [Pic68]. Any d.f. in the domain of attraction $D(\Lambda)$ of the double exponential extreme value distribution belongs to $R_{-\infty}$ [Res87]. For example, the exponential d.f. $\overline{F}(t) = e^{-\lambda t}$ yields $\mathbf{E}X_{(n)}/n \sim \ln n/(\lambda n)$.

(5). See [Dow90b]. E^1 is the class of random variables dominated in convex ordering [Ros83, Sto83] by some exponentially distributed variate. E^1 is a very large class of random variables, including all those [Dow90b] that are Coxian or of Phase (*PH*) type, those with bounded mean residual life, those that are New Better Than Used in Expectation (*NBUE*), and all subclasses [Ros83] of the latter, such as Increasing Failure Rate (*IFR*) and Increasing Likelihood Ratio (*ILR*) variates.

(6). E^β is the class of variates X^β for some X in E^1. An example is the Weibull $\overline{F}(t) = \exp(-\lambda t^{1/\beta})$.

Table 1. Summary of $\mathbf{E}X_{(n)}/n$ Bounds

Often one needs to know the asymptotic behavior of integrals of regularly varying functions; for example, when estimating the moments of regularly varying r.v.s. This integral behavior is given by the celebrated Karamata's Theorem, only one direction of which will be cited here.

THEOREM (Karamata's Theorem; direct half) [Kar30, Bin87]. If $f \in R_\rho$ and f is locally bounded in $[a, \infty)$, then

(i) For any $\sigma \geq -(\rho + 1)$,

$$\frac{x^{\sigma+1} f(x)}{\int_a^x t^\sigma f(t)\, dt} \to \sigma + \rho + 1 \qquad (x \to \infty)\ . \tag{2.1.3}$$

If $\rho = \infty$, the limit on the right is interpreted as ∞.

(ii) For any $\sigma < -(\rho + 1)$, (and for $\sigma = -(\rho + 1)$ if $\int_a^\infty t^{-(\rho+1)} f(t)\, dt < \infty$) then

$$\int_x^\infty t^\sigma f(t)\, dt$$

converges for $x \in [a, \infty)$ and

$$\frac{x^{\sigma+1} f(x)}{\int_x^{\infty} t^{\sigma} f(t)\, dt} \quad \to \quad -(\sigma + \rho + 1) \qquad (x \to \infty) .$$

If $\rho = -\infty$, the limit on the right is interpreted as ∞. □

Karamata's Theorem finds application in probability theory by describing the behavior of moments of random variables with well-behaved tails. The first result says that any regularly varying r.v. of index $-\alpha$ has finite moments of all orders in $[0, \alpha)$:

THEOREM. Suppose that X is an r.v. with $X \in R_{-\alpha}$ for some $\alpha > 0$. Then

$$\forall \beta < \alpha \qquad \mathbf{E}[X^{\beta}] < \infty . \tag{2.1.4}$$

Proof: Apply Karamata's Theorem (2.1.3) with f set to $\overline{F}$, a set to 0, ρ set to $-\alpha$ and σ set to $\beta - 1$. Since $\beta < \alpha \Rightarrow \sigma < -(\rho + 1)$, part (ii) applies and yields

$$\int_x^{\infty} t^{\beta - 1} \overline{F}(t) dt < \infty$$

for all $x \ge 0$. By (1.3), this implies $\mathbf{E}[X^{\beta}] < \infty$. □

Another useful consequence of this theorem in probability theory describes the behavior of the mean residual life of a regularly varying r.v. Let $X =_d F$. The *residual life at time t* is the random variable $X_t := X - t \mid X > t$ having d.f. $1 - \overline{F}(x+t)/\overline{F}(t)$. If we assume that $\mathbf{E}X < \infty$, then the expectation of the mean residual life at t

$$\mathbf{E}X_t = \frac{\int_t^{\infty} \overline{F}(u)\, du}{\overline{F}(t)}$$

is called the *mean residual life function* of X.

THEOREM. Suppose $X \in R_{-\alpha}$ for some $\alpha > 1$ (so that $\mathbf{E}X < \infty$). Then

$$\mathbf{E}X_t \sim \frac{t}{\alpha - 1} \qquad (t \to \infty) . \tag{2.1.5}$$

If $\alpha = \infty$, this is interpreted as saying $\mathbf{E}X_t = o(t)$.

Proof: Apply Karamata's Theorem (2.1.3), part (ii), with a set to 0, f set to $\overline{F}$, ρ set to $-\alpha$ and σ set to 0. □

We see from the above that for distributions with regularly decaying tails, the mean residual life grows linearly with time. In the case of rapidly decaying tails ($\overline{F} \in R_{-\infty}$), the mean residual life grows more slowly than linearly.

2.2 Results about Extremes

Note that if $X_1, \cdots, X_n$ are i.i.d., $=_d F(t)$, then $X_{(n)} =_d F^n(t)$. It is an easy consequence [Dow90a] that for all p, $\mathbf{E}X^p < \infty \Leftrightarrow \mathbf{E}X_{(n)}^p < \infty$.

THEOREM [Dow90a]. Suppose X_i are i.i.d. $=_d F$. Then

$$\mathbf{E}X_1 < \infty \;\Rightarrow\; \frac{\mathbf{E}X_{(n)}}{n} \to 0 \qquad (n \to \infty) \tag{2.2.1}$$

Proof:

$$\mathbf{E}X_{(n)} = \int_0^{\infty} t\, dF^n(t) = n \int_0^{\infty} t F^{n-1}(t)\; dF(t)$$

It is enough to show

$$I_n := \int_0^\infty tF^{n-1}(t)\,dF(t) = o(1) \qquad (n \to \infty)$$

Break up the integral into two parts by decomposing the range of integration:

$$I_n = H_n \quad + \quad T_n$$

where

$$H_n := \int_0^{\ln n} tF^{n-1}(t)\,dF(t) \quad , \qquad T_n := \int_{\ln n}^\infty tF^{n-1}(t)\,dF(t) \quad .$$

First, the head integral tends to zero since

$$H_n \le \ln n \cdot \int_0^{\ln n} F^{n-1}(t)\,dF(t) = \ln n \cdot \int_0^{F(\ln n)} u^{n-1} du = \ln n \cdot \frac{F^n(\ln n)}{n} \to 0 \ .$$

Finally, for the tail integral we have

$$T_n \le \int_{\ln n}^\infty t\,dF(t) \to 0$$

since

$$\int_0^\infty t\,dF(t) = \mathbf{E}X_1 < \infty \qquad \square$$

COROLLARY. Suppose X_i are i.i.d. $=_d F$. Then for all $p \ge 1$

$$\mathbf{E}[X_1^p] < \infty \ \Rightarrow \ \mathbf{E}[{X_{(n)}}^p] = o(n) \qquad (n \to \infty) \tag{2.2.2}$$

Proof: Define $Y_i := X_i^p$. Then $Y_1, \cdots, Y_n$ are i.i.d. with $\mathbf{E}Y_1 < \infty$. So by Theorem 2.2.1, $\mathbf{E}Y_{(n)} = o(n)$. However

$$Y_{(n)} = \max(Y_1, \cdots, Y_n) = \max(X_1^p, \cdots, X_n^p) = \max(X_1, \cdots, X_n)^p = X_{(n)}^p$$

and so $\mathbf{E}[X_{(n)}^p] = o(n)$. $\square$

COROLLARY. Suppose X_i are i.i.d. $=_d F$. Then

$$\mathbf{E}X_1^2 < \infty \ \Rightarrow \ \frac{\mathbf{E}X_{(n)}^2}{n^2} = o\left(\frac{\mathbf{E}X_{(n)}}{n}\right) \qquad (n \to \infty) \ . \tag{2.2.3}$$

Proof: By the previous result, $\mathbf{E}X_{(n)}^2/n = o(1)$ as $n \to \infty$. Since $\mathbf{E}X_{(n)} \ge \mathbf{E}X = \Theta(1)$, it follows that $\mathbf{E}X_{(n)}^2/n = o(\mathbf{E}X_{(n)})$ and division by n completes the argument. $\square$

The next result has an Abelian flavor.

THEOREM. Suppose X_i are i.i.d. $=_d F$. Suppose that F has infinite support, so that $\mathbf{E}X_{(n)} \to \infty$. If

(i) $\quad \mathbf{E}X_1 < \infty$

(ii) $\quad f(t) \le t$

(iii) $\quad f(t) = o(t) \qquad (t \to \infty)$

then

$$\mathbf{E}[fX_{(n)}] = o(\mathbf{E}X_{(n)}) \qquad (n \to \infty) \tag{2.2.4}$$

Proof: Write $\mu := \mathbf{E}X_1$. By hypothesis (iii),

$$\forall \varepsilon > 0 \ \ \exists N(\varepsilon) \ \ \forall t > N(\varepsilon) \ \ f(t) \le \varepsilon t$$

Without loss of generality we may assume $N(\varepsilon) > \varepsilon\mu$. Define

$$N'(\varepsilon) := \max\left(N(\varepsilon/2),\ \frac{\left| \ln \dfrac{\mu\varepsilon/2}{N(\varepsilon/2)} \right|}{\left| \ln F(N(\varepsilon/2)) \right|} \right) \ .$$

Then for all $n > N'(\varepsilon)$ we have

$$n > \frac{\ln[\tfrac{1}{2}\mu\varepsilon/N(\varepsilon/2)]}{\ln F(N(\varepsilon/2))} \quad \Rightarrow \quad \ln F(N(\varepsilon/2))^n < \ln[\tfrac{1}{2}\mu\varepsilon/N(\varepsilon/2)]$$

which implies

$$N(\varepsilon/2)\cdot F(N(\varepsilon/2))^n < \frac{\varepsilon\mu}{2} \le \frac{\varepsilon}{2}\mathbf{E}X_{(n)} \ . \tag{2.2.5}$$

Also

$$\forall t \ge N(\varepsilon/2) \quad f(t) \le \frac{\varepsilon}{2}t \ . \tag{2.2.6}$$

We can express the expectation as a sum of two integrals:

$$\mathbf{E}[fX_{(n)}] = \int_0^\infty f(t)\,dF^n(t) = \int_0^{N(\varepsilon/2)} f(t)\,dF^n(t) + \int_{N(\varepsilon/2)}^\infty f(t)\,dF^n(t)$$

Using hypotheses (ii) in the first integral and equation (2.2.6) in the second, we get

$$\mathbf{E}[fX_{(n)}] \le \int_0^{N(\varepsilon/2)} t\,dF^n(t) + \frac{\varepsilon}{2}\int_{N(\varepsilon/2)}^\infty t\,dF^n(t) \le N(\varepsilon/2)\int_0^{N(\varepsilon/2)} dF^n(t) + \frac{\varepsilon}{2}\mathbf{E}X_{(n)}$$

$$\le N(\varepsilon/2)\cdot F^n(N(\varepsilon/2)) + \frac{\varepsilon}{2}\mathbf{E}X_{(n)}$$

Using equation (2.2.5) we have that for all $n > N'(\varepsilon)$

$$\mathbf{E}[fX_{(n)}] \le \frac{\varepsilon}{2}\mathbf{E}X_{(n)} + \frac{\varepsilon}{2}\mathbf{E}X_{(n)} = \varepsilon\mathbf{E}X_{(n)} \ .$$

This last assertion shows that

$$\forall\varepsilon > 0 \ \ \exists N'(\varepsilon) \ \ \forall n > N'(\varepsilon) \quad \mathbf{E}[fX_{(n)}] \le \varepsilon\mathbf{E}X_{(n)} \ ,$$

that is, $\mathbf{E}[fX_{(n)}] = o(\mathbf{E}X_{(n)})$. □

The characteristic value, given n, is the value of the random variable that will be exceeded with probability at most $1/n$:

DEFINITION. For a random variable $X =_d F$, the *characteristic maximum function* $c_X(n)$ is defined as the $1-n^{-1}$ quantile point of the d.f.:

$$c_X(n) := \inf\{x \mid \overline{F}(x) \le n^{-1}\} \ . \tag{2.2.7}$$

We now cite a result that connects the behavior of the expected extreme with that of the characteristic maximum and the mean residual life:

THEOREM [Dow91, Dow93]. Let X be any r.v. with finite first moment. Then if $c_X(n)$ is the characteristic maximum function and $\mathbf{E}X_t$ is the mean residual life function, then for all $n \ge 1$

$$(1 - e^{-1})\left[c_X(n) + \mathbf{E}X_{c_X(n)}\right] < \mathbf{E}X_{(n)} \le \left[c_X(n) + \mathbf{E}X_{c_X(n)}\right] , \tag{2.2.8}$$

that is,

$$\mathbf{E}X_{(n)} = \Theta\left[c_X(n) + \mathbf{E}X_{c_X(n)}\right] \qquad (n \to \infty) \ . \qquad \square$$

The above upper/lower bound on $\mathbf{E}X_{(n)}$ can be immediately applied to characterize the order of growth of moments of $X_{(n)}$, provided X is regularly varying. Although the expectation of the extreme is not in general equal to the characteristic maximum, it is close:

THEOREM. Assume $X \in R_{-\alpha}$ for some $\alpha > 1$. Then for all $\beta < \alpha$:

$$\mathbf{E}[X_{(n)}^\beta] = \Theta(c_X(n)^\beta) \ . \tag{2.2.9}$$

Proof: It is easy to see that

$$X^{\beta} =_d F(t^{1/\beta}) \ .$$

From this d.f., and since since $X \in R_{-\alpha}$, it follows from (2.1.2) that $X^{\beta} \in R_{-\alpha/\beta}$. Also from this d.f. it follows that

$$c_{X^{\beta}}(n) = c_X(n)^{\beta} \ .$$

Since X^{β} is regularly varying, Theorem (2.1.5) yields

$$\mathbf{E}X_t^{\beta} \sim \frac{t}{\alpha/\beta - 1} \ .$$

Using the two foregoing facts, Theorem (2.2.8) now applies to X^{β} to yield

$$\mathbf{E}[X_{(n)}^{\beta}] = \Theta\left[c_X(n)^{\beta} + (\alpha/\beta - 1)^{-1} c_X(n)^{\beta}\right] = \Theta(c_X(n)^{\beta}) \qquad (n \to \infty) \ . \square$$

A consequence of the above is that for regularly varying r.v.'s, the βth moment of $X_{(n)}$ is approximately the βth power of its first moment: $\mathbf{E}[X_{(n)}^{\beta}] = \Theta((\mathbf{E}X_{(n)})^{\beta})$. The following consequence is needed in the sequel. It complements Corollary (2.2.3) in the case of regularly varying $\bar{F}$.

COROLLARY. If $X \in R_{-\alpha}$, for some $\alpha > 1$, then

$$\forall\, \beta \in (1, \alpha) \quad \frac{\mathbf{E}[X_{(n)}^{\beta}]}{n^{\beta}} = o\left(\frac{\mathbf{E}X_{(n)}}{n}\right) \qquad (n \to \infty) \ . \tag{2.2.10}$$

Proof: From Theorem (2.2.9), we have two equations:

$$\frac{\mathbf{E}[X_{(n)}^{\beta}]}{n^{\beta}} = \Theta\left(\left[\frac{c_X(n)}{n}\right]^{\beta}\right)$$

and

$$\frac{\mathbf{E}[X_{(n)}]}{n} = \Theta\left(\frac{c_X(n)}{n}\right) \ .$$

By Theorem (2.2.1) and the second equation, $c_X(n)/n = o(1)$. This observation and the fact that $\beta > 1$ yield

$$\left[\frac{c_X(n)}{n}\right]^{\beta} = o\left(\frac{c_X(n)}{n}\right) \ .$$

Since the right side of the first equation is o of the right side of the second equation, the result follows. $\square$

2.3 Laplace Transform Bounds

For a non-negative random variable $X =_d F(t)$, its *Laplace-Stieltjes Transform* $\tilde{F}(x)$ is defined as

$$\tilde{F}(x) := \mathbf{E}[e^{-xX}] = \int_0^{\infty} e^{-xt}\, dF(t) \ . \tag{2.3.1}$$

The moment $\mathbf{E}X^p$ is finite $\Leftrightarrow$ a finite limit $\tilde{F}^{(p)}(0+)$ exists [Fel66, XIII.2].

Various upper and lower bounds on $\tilde{F}(x)$ will be of use in the sequel. All moments $\mathbf{E}X^p$ exist and are either finite or infinite; in the infinite case the results below have the obvious degenerate (and trivial) meanings.

THEOREM. For any d.f. F of a non-negative variate, and for all $x \geq 0$

$$\tilde{F}(x) \geq e^{-x\mathbf{E}X} \ . \tag{2.3.2}$$

Proof: $\tilde{F}(x) = \mathbf{E}[e^{-xX}] \geq e^{-x\mathbf{E}X}$ by Jensen's inequality [Chu74]. $\square$

COROLLARY. For any d.f. F of a non-negative variate, and for all $x \geq 0$

$$\tilde{F}\left(\frac{x}{m}\right)^m \geq e^{-x\mathbf{E}X} \ . \qquad \square \tag{2.3.3}$$

THEOREM. Let $X =_d F$ have finite first moment. Then for all x in $[0, 1/\mathbf{E}X]$:

$$\tilde{F}(x) \leq e^{-x\mathbf{E}X/2} . \tag{2.3.4}$$

Proof: The function $\tilde{F}(x)$ is convex decreasing on $[0, \infty)$ with $\tilde{F}(0) = 1$ and $\tilde{F}(\infty) = 0$ [Wid71]. At $x = 0$ it has tangent line $1 - x\mathbf{E}X$. Therefore the line $1 - \frac{1}{2}x\mathbf{E}X$ passes through this tangent point with larger (negative) slope and so is a secant line, intersecting the curve $\tilde{F}(x)$ at points $x = 0$ and $x = \theta > 0$. For every x in $[0, \theta]$, we therefore have

$$\tilde{F}(x) \leq 1 - \tfrac{1}{2}x\mathbf{E}X \leq e^{-\frac{1}{2}x\mathbf{E}X} . \tag{†}$$

Since the tangent line $1 - x\mathbf{E}X$ intercepts the x-axis at $1/\mathbf{E}X$, it must be the case that $1/\mathbf{E}X < \theta$, so for all x to the left of $1/\mathbf{E}X$, (†) holds. □

COROLLARY. Let $X =_d F$ have finite first moment. Then for all x with $0 \leq x/m \leq 1/\mathbf{E}X$:

$$\tilde{F}(\frac{x}{m})^m \leq e^{-x\mathbf{E}X/2} . \qquad \square \tag{2.3.5}$$

LEMMA. If $\mathbf{E}X < \infty$ then for all $x \geq 0$

$$\tilde{F}(x) = 1 - x\mathbf{E}X + x\int_0^\infty \overline{F}(t)(1 - e^{-xt})\, dt . \tag{2.3.6}$$

Proof: Integration of (2.3.1) by parts yields

$$\tilde{F}(x) = 1 - x\int_0^\infty \overline{F}(t)e^{-xt}\, dt ,$$

and using the fact that

$$\mathbf{E}X = \int_0^\infty \overline{F}(t)\, dt$$

in the previous equation yields (2.3.6). □

THEOREM. Let $X =_d F$ have finite first moment. Then for all $x \geq 0$

$$\tilde{F}(x) \leq \exp[-x\mathbf{E}X]\cdot\exp[x\int_0^\infty \overline{F}(t)(1 - e^{-xt})\, dt] . \tag{2.3.7}$$

Proof: One easily checks that for all non-negative x

$$x\mathbf{E}X \geq x\int_0^\infty \overline{F}(t)(1 - e^{-xt})\, dt$$

Thus Lemma (2.3.6) and the fact that $1 - w \leq e^{-w}$ for non-negative w imply the result. □

COROLLARY. Let $X =_d F$ have finite first moment. Then for all $x \geq 0$

$$\tilde{F}(\frac{x}{m})^m \leq \exp[-x\mathbf{E}X]\cdot\exp[x\int_0^\infty \overline{F}(t)(1 - e^{-xt/m})\, dt] . \qquad \square \tag{2.3.8}$$

REMARK. Taken together, Corollary (2.3.3) and (2.3.8) imply the following pointwise limit for each fixed x, assuming a finite first moment:

$$\lim_{m \to \infty} \tilde{F}(\frac{x}{m})^m = e^{-x\mathbf{E}X} .$$

The right hand side is the transform of the d.f. concentrated at $\mathbf{E}X$, and the left hand side is the transform of $(X_1 + \cdots + X_m)/m$. Therefore, using the fact that Laplace transforms are unique, we have derived the weak law of large numbers for non-negative random variables, due to Khinchine in this form [Fel66, XIII.3]. □

COROLLARY. Let $X =_d F$ be a random variable with $\mathbf{E}[X^{1+\nu}] < \infty$ for some $1 \ge \nu > 0$. Then for all $x \ge 0$

$$\tilde{F}(\frac{x}{m})^m \le e^{-x\mathbf{E}X} e^{x^{1+\nu} m^{-\nu} \mathbf{E}[X^{1+\nu}]/(1+\nu)} \ . \tag{2.3.9}$$

Proof: We make use of the inequality

$$\forall\, w \ge 0 \ \forall\, \nu \in [0,1] \quad 1 - e^{-w} \le w^{\nu} \ . \tag{2.3.10}$$

Using this inside the integral in (2.3.8) results in

$$x\int_0^\infty \bar{F}(t)(1 - e^{-xt/m})\, dt \le \frac{x^{1+\nu}}{m^{\nu}} \int_0^\infty \bar{F}(t) t^{\nu}\, dt \ .$$

The latter integral is $(1+\nu)^{-1}\mathbf{E}[X^{1+\nu}]$ by (1.3). □

2.4 The Truncated Random Variable

Given a variate $X =_d F$ we can define for each $z \ge 0$ that is in the support of F a new *truncated* variate $Y(z)$ that represents the random variate X conditioned on the event $[X \le z]$:

$$Y(z) = X \mid X \le z \tag{2.4.1}$$

Let $Y(z) =_d G_z(t)$. Then

$$G_z(t) = \begin{cases} F(t)/F(z) & 0 \le t \le z \\ 1 & t \ge z \end{cases} \tag{2.4.2}$$

The pth moment of the truncated variate is given by

$$\mathbf{E}Y^p(z) = \frac{\int_0^z t^p \ dF(t)}{F(z)} = \frac{\mu_p(z)}{F(z)} \ , \tag{2.4.3}$$

and in particular

$$\mathbf{E}Y(z) = \mu(z)/F(z) \ .$$

Notice that since $Y(z)$ has bounded support, all moments of $Y(z)$ are finite for finite z. The random variable $Y(z)$ has the Laplace-Stieltjes transform

$$\tilde{G}_z(x) = \mathbf{E}[e^{-xY(z)}] = \int_0^z e^{-xt} \frac{dF(t)}{F(z)} \tag{2.4.4}$$

The following are general properties of the moments $\mathbf{E}Y^p(z)$ valid for any d.f. F.

THEOREM. For any d.f. F of a non-negative variate, let $Y(z)$ be the truncated variate. Then for any z in the support of F and for any $p > 0$

$$\mathbf{E}Y(z)^p = \frac{\mu_p(z)}{F(z)} \uparrow z \tag{2.4.5}$$

Proof: By parts we derive

$$\mathbf{E}Y(z)^p = z^p - \frac{\int_0^z F(t) dt^p}{F(z)} \ . \tag{2.4.6}$$

Two facts are needed for the sequel. By monotonicity of d.f. F

$$\forall z, \varepsilon \ge 0 \quad F(z) \le F(z+\varepsilon) \ . \tag{2.4.7}$$

Also by monotone properties of integration

$$\forall z, \varepsilon \ge 0 \quad \int_z^{z+\varepsilon} F(t)dt^p \le [(z+\varepsilon)^p - z^p]\cdot F(z+\varepsilon) \ . \tag{2.4.8}$$

Let z and ε be arbitrary and non-negative, with both z and $z+\varepsilon$ in the support of F. From (2.4.7) we obtain

$$0 \ge \frac{\int_0^z F(t)\,dt^p}{F(z+\varepsilon)} - \frac{\int_0^z F(t)\,dt^p}{F(z)} \tag{2.4.9}$$

From (2.4.8) we obtain

$$(z+\varepsilon)^p - z^p \ge \frac{\int_z^{z+\varepsilon} F(t)\,dt^p}{F(z+\varepsilon)} \ ,$$

and adding this to (2.4.9) yields

$$(z+\varepsilon)^p - z^p \ge \frac{\int_0^{z+\varepsilon} F(t)\,dt^p}{F(z+\varepsilon)} - \frac{\int_0^z F(t)\,dt^p}{F(z)} \tag{2.4.10}$$

Rearranging (2.4.10) results in:

$$(z+\varepsilon)^p - \frac{\int_0^{z+\varepsilon} F(t)\,dt^p}{F(z+\varepsilon)} \quad \ge \quad z^p - \frac{\int_0^z F(t)\,dt^p}{F(z)}$$

or using identity (2.4.6)

$$\mathbf{E}Y(z+\varepsilon)^p \ge \mathbf{E}Y(z)^p \quad .$$

This last result is the desired monotonicity. □

REMARK. It is also possible to show using the mean value theorem [Wid61] that $\mathbf{E}Y^p(0^+) = 0$, that $\mathbf{E}Y^p(z) = o(z^p)$ as $z \to 0^+$, and that $\mathbf{E}Y^p(z) \le z^p$ for all $z > 0$.

Obviously as $z \to \infty$, $\mathbf{E}Y(z)^p \to \mathbf{E}X^p$.

3. Integral Representation for $\mathbf{E}R_n$

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$. Then

$$\mathbf{E}R_n \ = \ \int_0^\infty z \int_0^\infty e^{-zx}[\tilde{G}_z(x)]^{n-1}\,dx\,dF^n(z) \tag{3.1}$$

where $\tilde{G}_z$ is the Laplace-Stieltjes transform of the truncated variate $Y(z) := X \mid X \le z$:

$$\tilde{G}_z(x) = \frac{\int_0^z e^{-xt}\,dF(t)}{F(z)} \ .$$

Proof: Let $X_1, X_2, \cdots, X_n$ be a sample of n independent random variables with distribution F, and let their ordered values be the order statistics

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n-1)} \le X_{(n)} \ .$$

Conditioned on $X_{(n)} = z$, the remaining $n-1$ variates of the sample are independent and identically distributed random variables, each having the d.f. G_z of $Y(z)$, defined in (2.4.2). Call these $n-1$ i.i.d. variates

$$Y(z)_1, Y(z)_2, \cdots, Y(z)_{n-1} \ .$$

Define their sum as

$$Z_{n-1}(z) := Y(z)_1 + Y(z)_2 + \cdots + Y(z)_{n-1} .$$

Then we have the following conditional expectation

$$\mathbf{E}[R_n \mid X_{(n)} = z] = \mathbf{E}\left[\frac{z}{z + Z_{n-1}(z)}\right] .$$

Removing the conditioning, using the fact that $X_{(n)} =_d F^n(z)$, results in

$$\mathbf{E}R_n = \mathbf{E}\left[\int_0^\infty \frac{z}{z + Z_{n-1}(z)}\, dF^n(z)\right]$$

In the above integral, use the fact that

$$\frac{1}{w} = \int_0^\infty e^{-wx}\, dx$$

to obtain

$$\mathbf{E}R_n = \mathbf{E}\left[\int_0^\infty z \int_0^\infty e^{-x[z + Z_{n-1}(z)]}\, dx\, dF^n(z)\right] .$$

Since all the iterated integrals in the above expression are convergent for all parameter values, and since all integrands are non-negative, we may employ Fubini's Theorem [Fel66, IV.2] to reorder the integrals, resulting in

$$\mathbf{E}R_n = \int_0^\infty z \int_0^\infty e^{-xz}\, \mathbf{E}[e^{-xZ_{n-1}(z)}]\, dx\, dF^n(z) . \tag{3.2}$$

Now because $Z_{n-1}(z)$ is a sum of i.i.d. variates, the convolution theorem holds for Laplace transforms and

$$\mathbf{E}[e^{-xZ_{n-1}(z)}] = \mathbf{E}[e^{-xY(z)}]^{n-1} = [\tilde{G}_z(x)]^{n-1} .$$

Putting this into (3.2) yields the result (3.1). □

For specific d.f.s, direct asymptotic expansion of the integral form (3.1) is often possible.

EXAMPLE. Consider the uniform distribution on $[0, 1]$, where $F(t) = t$. Then

$$\tilde{G}_z(x) = \frac{1 - e^{-xz}}{xz} ,$$

and (3.1) becomes

$$\mathbf{E}R_n = \int_0^1 z \int_0^\infty e^{-xz} \left(\frac{1 - e^{-xz}}{xz}\right)^{n-1} dx\, dz^n = \int_0^\infty e^{-w} \left(\frac{1 - e^{-w}}{w}\right)^{n-1} dw ,$$

after a change of variable. A further integration by parts yields a simpler looking form:

$$\mathbf{E}R_n = \frac{n-1}{n} \int_0^\infty \frac{(1 - e^{-w})^n}{w^n}\, dw .$$

The function $(1 - e^{-w})/w$ decreases in w from its maximum at $w = 0$. This integral is a classic Laplace integral [Olv74], and asymptotic expansion for large n proceeds as follows. First, we may replace the upper limit of integration by 1, as the contribution to the integral over the range $[1, \infty)$ is easily seen to be subdominant. So:

$$\mathbf{E}R_n = \frac{n-1}{n} \int_0^1 \frac{(1 - e^{-w})^n}{w^n}\, dw + \varpi(n) ,$$

where $\varpi(n)$ represents a subdominant term. Use the first few terms in the Taylor expansion of $(1 - e^{-w})/w$ to force the integrand to be of the form e^{-un}; this dictates a choice of u such that

$$e^{-u} = 1 - \frac{w}{2} + \frac{w^2}{6} .$$

This amounts to the change (ignoring any higher terms) given by

$$u \sim \frac{w}{2} - \frac{w^2}{24} .$$

Now solve for w in terms of u ($w = 6 + 6\sqrt{1 - 2/3u}$) and make this change of variable in the integral to get

$$\mathbf{E}R_n = \frac{n-1}{n} \int_0^b e^{-un} \frac{2}{\sqrt{1 - 2/3u}} \, du + \omega(n)$$

where $b = 11/24$.

Apply the standard series expansion to the square root to get

$$\mathbf{E}R_n = \frac{2(n-1)}{n} \int_0^b e^{-un} \left[1 + \frac{u}{3} + O(u^2) \right] du + \omega(n) = \frac{2(n-1)}{n^2} \int_0^{bn} e^{-z} \left[1 + \frac{z}{3n} + O(z^2/n^2) \right] dz + \omega(n) .$$

Setting $b = \infty$ does not affect the approximation except in exponentially small terms [Olv74], and so the integrals are all elementary. Evaluating them yields

$$\mathbf{E}R_n = \frac{2(n-1)}{n^2} \left[1 + \frac{1}{3n} + O(n^{-2}) \right] \qquad (n \to \infty) .$$

This agrees in first term with the estimate first derived in [Cof85], but provides higher order terms. □

4. Lower Bound

To find an asymptotic lower bound for the integral (3.1), we will bound from below the Laplace transform $\tilde{G}_z(y/n)^n$, using the weak law lower bound from Section 2.3. This is essentially application of Jensen's equality followed by expansion in terms of moments.

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$. Suppose either that

(i) $\bar{F} \in R_{-\alpha}$ for some $\alpha > 1$; or else

(ii) F has finite second moment.

Then

$$\mathbf{E}R_n \geq \frac{\mathbf{E}X_{(n)}}{\mathbf{E}S_n} \left[1 + o(1) \right] \qquad (n \to \infty) \tag{4.1}$$

Proof: Make the change $x := y/n$ in (3.1) to get

$$\mathbf{E}R_n = \frac{1}{n} \int_0^\infty z \int_0^\infty e^{-yz/n} [\tilde{G}_z(y/n)]^{n-1} \, dy \, dF^n(z) .$$

Let μ be $\mathbf{E}X_1$. By Corollary (2.3.3), the fact (2.4.5) that $\mathbf{E}Y(z) \uparrow z$, and the observation $\mathbf{E}Y(\infty) = \mu$, we have

$$\tilde{G}_z(y/n)^{n-1} \geq \tilde{G}_z(y/n)^n \geq e^{-y\mathbf{E}Y(z)} \geq e^{-y\mu}$$

Using this lower bound in the above integral gives

$$\mathbf{E}R_n \geq \frac{1}{n} \int_0^\infty z \int_0^\infty e^{-yz/n} e^{-y\mu} \, dy \, dF^n(z) = \frac{1}{n\mu} \int_0^\infty \frac{z}{1 + z/(n\mu)} \, dF^n(z) .$$

Since $(1 + w)^{-1} \geq e^{-w}$, we have the lower bound

$$\mathbf{E}R_n \geq \frac{1}{n\mu} \int_0^\infty z e^{-z/(n\mu)} \, dF^n(z) . \tag{4.2}$$

The proof now divides into two parts, depending on whether hypothesis (i) or (ii) obtains.

(i). Assume $\bar{F} \in R_{-\alpha}$ for $\alpha > 1$. Pick a $\nu \in (0,1)$ such that $1 + \nu < \alpha$. Then by Theorem (2.1.4), $\mathbf{E}X_1^{1+\nu} < \infty$. Now

$$\forall w \geq 0 \quad e^{-w} \geq 1 - w^{\nu} \ ,$$

and applying this to (4.2) yields

$$\mathbf{E}R_n \ \geq \ \frac{1}{n\mu}\int_0^{\infty} z\left[1 - \frac{z^{\nu}}{n^{\nu}\mu^{\nu}}\right] dF^n(z) \ = \ \frac{\mathbf{E}X_{(n)}}{n\mu} \ - \ \frac{\mathbf{E}[X_{(n)}^{1+\nu}]}{n^{1+\nu}\mu^{1+\nu}} \ . \tag{4.3}$$

By Corollary (2.2.10), since $\overline{F} \in R_{-\alpha}$ with $1+\nu \in (1, \alpha)$, we have

$$\frac{\mathbf{E}[X_{(n)}^{1+\nu}]}{n^{1+\nu}} \ = \ o\left[\frac{\mathbf{E}X_{(n)}}{n}\right] \qquad (n \to \infty) \ .$$

Using this in (4.3) shows that the second term is asymptotically smaller than the first, so that we conclude

$$\mathbf{E}R_n \geq \frac{\mathbf{E}X_{(n)}}{\mu n}\left[1 + o(1)\right] \qquad (n \to \infty) \ . \tag{4.4}$$

Since $\mathbf{E}S_n = \mu n$, we have the desired result.

(ii). Assume X_1 has a finite second moment. Use the inequality

$$\forall w \geq 0 \quad e^{-w} \geq 1 - w$$

in equation (4.2) to obtain

$$\mathbf{E}R_n \ \geq \ \frac{\mathbf{E}X_{(n)}}{n\mu} \ - \ \frac{\mathbf{E}[X_{(n)}^2]}{n^2\mu^2} \ . \tag{4.5}$$

Now by applying Corollary (2.2.3), we conclude that the second term is asymptotically smaller than the first, and so (4.4) follows once again. □

5. Upper Bound

We proceed now to the upper bound. The approach is to use the weak law upper bounds on $\tilde{G}_z(y/n)^n$ from Section 2.3.

First we show that contributions to $\mathbf{E}R_n$ from the integral (3.1) for z near the origin are exponentially small and may be ignored.

Call a function $a(n)$ *subdominant* if, for every $k \geq 0$ $a(n) = o(n^{-k})$ as $n \to \infty$. For convenience, we denote the class of subdominant functions by $\omega(n)$. Any function in this class will also be denoted by $\omega(n)$. Such functions are exponentially small and play the role of zeros in any asymptotic expansion.

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$. Let $\delta \geq 0$ be any point in the support of F, so that $F(\delta) < 1$. Then

$$\mathbf{E}R_n \ = \ \int_{\delta}^{\infty} z \int_0^{\infty} e^{-zx}[\tilde{G}_z(x)]^{n-1}\, dx\, dF^n(z) \ + \ \omega(n) \ . \tag{5.1}$$

Proof: Since $\tilde{G}_z(x) \downarrow x$ and $\tilde{G}_z(0) \leq \int_0^z dF(t)/F(z) = 1$, the part of the integral (3.1) near the origin is

$$\int_0^{\delta} z \int_0^{\infty} e^{-zx}[\tilde{G}_z(x)]^{n-1}\, dx\, dF^n(z) \ \leq \ \int_0^{\delta} z \int_0^{\infty} e^{-zx}\, dx\, dF^n(z) \ = \ \int_0^{\delta} dF^n(z) = F^n(\delta) - F^n(0) = \omega(n) \quad \square$$

The next step eliminates the inner integral in (5.1) at the cost of $o(1)$ terms.

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$, and that $\mathbf{E}X_1^{1+\nu} < \infty$ for some $\nu > 0$. Then for any δ with $F(\delta) < 1$

$$\mathbf{E}R_n \ \leq \ \frac{1}{n}\int_{\delta}^{\infty} z \frac{1}{z/n + \mu(z)/F(z)} \cdot dF^n(z)\left[1 \ + \ o(1)\right] \qquad (n \to \infty) \tag{5.2}$$

Proof: From Theorem (5.1), it is sufficient to show that the bound in (5.2) dominates

$$J_n := \int_\delta^\infty z \int_0^\infty e^{-zx} [\tilde{G}_z(x)]^{n-1} \, dx \, dF^n(z) \ .$$

Make the change $x := y/n$ to obtain

$$J_n := \frac{1}{n} \int_\delta^\infty z \int_0^\infty e^{-zy/n} [\tilde{G}_z(y/n)]^{n-1} \, dy \, dF^n(z) \ .$$

Break up the integral into two integrals, denoted H_n and T_n, by splitting the range of the inner integral as follows:

$$J_n := H_n \ + \ T_n$$

where

$$H_n := \frac{1}{n} \int_\delta^\infty z \int_0^{r(n)} e^{-zy/n} \tilde{G}_z(y/n)^{n-1} \, dy \, dF^n(z) \ , \tag{5.3a}$$

and

$$T_n := \frac{1}{n} \int_\delta^\infty z \int_{r(n)}^\infty e^{-zy/n} \tilde{G}_z(y/n)^{n-1} \, dy \, dF^n(z) \ , \tag{5.3b}$$

and where $r(n)$ is chosen arbitrarily subject to the following constraints

$$r(n) = o(n^{\nu/(1+\nu)}) \qquad (n \to \infty) \tag{5.4a}$$

$$e^{-r(n)} = \omega(n) \qquad (n \to \infty) \ . \tag{5.4b}$$

For example the choice $r(n) = n^\beta$ for any β such that $0 < \beta < \nu/(1+\nu)$ will meet the constraints.

Below we show that T_n is subdominant and H_n has the bound in (5.2), completing the proof for J_n.

Tail integral T_n. Assume $y \geq r(n)$. Then

$$\frac{y}{n} \ \geq \ \frac{r(n)}{n} \ .$$

$\tilde{G}_z$ is a Laplace transform, so that $\tilde{G}_z(x) \downarrow x$, and we have

$$\tilde{G}_z(y/n)^{n-1} \ \leq \ \tilde{G}_z(r(n)/n)^{n-1} \ .$$

Consider the exponential factor in the integrand of T_n. We have by virtue of Theorem (2.3.2)

$$\tilde{G}_z(y/n)^{n-1} \leq \tilde{G}_z(r(n)/n)^{n-1} = \frac{\tilde{G}_z(r(n)/n)^n}{\tilde{G}_z(r(n)/n)}$$

$$\leq \ \frac{\tilde{G}_z(r(n)/n)^n}{e^{-r(n)\mathbf{E}Y(z)/n}} \ \leq \ [\tilde{G}_z(\, r(n)n^{-1/2}/\sqrt{n}\,)^{\sqrt{n}}]^{\sqrt{n}} \cdot e^{r(n)\mathbf{E}X/n} \ .$$

Equation (5.4a) implies that $r(n) = o(n)$ and hence that for sufficiently large n:

$$\frac{r(n)n^{-1/2}}{\sqrt{n}} \leq \frac{1}{\mathbf{E}X} \leq \frac{1}{\mathbf{E}Y(z)} \ , \tag{5.5}$$

since $\mathbf{E}Y(z) \uparrow z$. Therefore we can apply Corollary (2.3.5) to conclude that for all n sufficiently large:

$$\tilde{G}_z(y/n)^{n-1} \leq [e^{-\frac{1}{2} r(n)\mathbf{E}Y(z)/\sqrt{n}}]^{\sqrt{n}} \cdot e^{r(n)\mathbf{E}X/n} \ = \ e^{-\frac{1}{2}\mathbf{E}Y(z) r(n)} \cdot e^{r(n)\mathbf{E}X/n} \ .$$

Now for all sufficiently large n, $e^{r(n)\mathbf{E}X/n} \leq 2$ and $\mathbf{E}Y(z) \geq \mathbf{E}Y(\delta)$, so we finally have the bound

$$\tilde{G}_z(y/n)^{n-1} \leq 2e^{-\frac{1}{2}\mathbf{E}Y(\delta) r(n)} \tag{5.6}$$

which is valid for all $y \geq r(n)$ and all sufficiently large n. Putting this bound (5.6) into integral (5.3b) results in

$$T_n \le \frac{2}{n} e^{-\frac{1}{2}\mathbf{E}Y(\delta) r(n)} \int_\delta^\infty z \int_{r(n)}^\infty e^{-zy/n}\, dy\, dF^n(z) \;\le\; 2e^{-\frac{1}{2}\mathbf{E}Y(\delta) r(n)} \int_\delta^\infty e^{-zr(n)/n}\, dF^n(z) \;\le\; 2e^{-\frac{1}{2}\mathbf{E}Y(\delta)\cdot r(n)} \quad .$$

By (5.4b), this last bound is $\varpi(n)$, proving that

$$T_n = \varpi(n) \qquad (n \to \infty) \ . \tag{5.7}$$

Head integral H_n. Assume $y \le r(n)$. Then

$$\frac{y}{n} \le \frac{r(n)}{n} = o(n^{-1/(1+\nu)}) \ ,$$

by (5.4a). Hence by Corollary (2.3.9), writing $K(z)$ for $\mathbf{E}[Y(z)^{1+\nu}]/(1+\nu)$

$$\begin{aligned} \tilde{G}_z(y/n)^{n-1} &= [\tilde{G}_z(y/n)^n]^{(1-1/n)} \le [e^{-y\mathbf{E}Y(z)} \cdot e^{y^{1+\nu} n^{-\nu} K(z)}]^{(1-1/n)} \\ &\le e^{-y\mathbf{E}Y(z)} \cdot e^{y\mathbf{E}Y(z)/n} \cdot e^{K(z) y^{1+\nu}/n^\nu} \ . \end{aligned}$$

Now $y \le r(n)$, $K(z) \le K(\infty) = \mathbf{E}[X^{1+\nu}]/(1+\nu)$, and $\mathbf{E}Y(z) \le \mathbf{E}X$ so that we have the final bound

$$\tilde{G}_z(y/n)^{n-1} \le e^{-y\mathbf{E}Y(z)} \cdot e^{r(n)\mathbf{E}X/n} \cdot e^{K(\infty) r(n)^{1+\nu}/n^\nu} \ . \tag{5.8}$$

Putting this bound (5.8) into the integral (5.3a) results in

$$\begin{aligned} H_n &\le \frac{1}{n} e^{r(n)\mathbf{E}X/n} \cdot e^{K(\infty) r(n)^{1+\nu}/n^\nu} \int_\delta^\infty z \int_0^\infty e^{-zy/n} e^{-y\mathbf{E}Y(z)}\, dy\, dF^n(z) \\ &= \frac{1}{n} e^{r(n)\mathbf{E}X/n} \cdot e^{K(\infty) r(n)^{1+\nu}/n^\nu} \int_\delta^\infty \frac{z}{z/n + \mathbf{E}Y(z)}\, dF^n(z) \end{aligned}$$

From (5.4a) it follows that

$$r(n)^{1+\nu}/n^\nu = o(1) \qquad \text{and} \qquad r(n)/n = o(n^{-1/(1+\nu)}) \qquad (n \to \infty)$$

and this implies since $K(\infty) < \infty$ that

$$e^{r(n)\mathbf{E}X/n} \cdot e^{K(\infty) r(n)^{1+\nu}/n^\nu} = 1 + o(1) \qquad (n \to \infty) \ .$$

So we obtain the sought-for bound on H_n:

$$H_n \le \frac{1}{n} \int_\delta^\infty \frac{z}{z/n + \mathbf{E}Y(z)}\, dF^n(z) \left[1 + o(1)\right] \qquad (n \to \infty) \ . \tag{5.9}$$

Putting (5.9) and (5.7) together yields

$$J_n \le \frac{1}{n} \int_\delta^\infty \frac{z}{z/n + \mathbf{E}Y(z)}\, dF^n(z) \left[1 + o(1)\right] \qquad (n \to \infty) \ . \qquad \square$$

The final step in the argument shows that in (5.2) we may replace $\mu(z)/F(z)$ by its value at ∞, while introducing only further $o(1)$ terms.

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$, that F has unbounded support, and that $\mathbf{E}[X_1^{1+\nu}] < \infty$ for some $\nu > 0$. Then

$$n\mathbf{E}R_n \le \frac{\mathbf{E}X_{(n)}}{\mathbf{E}X_1} \left[1 + o(1)\right] \qquad (n \to \infty) \tag{5.10}$$

that is,

$$\mathbf{E}R_n \le \frac{\mathbf{E}X_{(n)}}{\mathbf{E}S_n} \left[1 + o(1)\right] \qquad (n \to \infty)$$

Proof: Write μ for $\mathbf{E}X_1$ in the sequel.

Define

$$I_n := \int_\delta^\infty z \frac{1}{z/n + \mu(z)/F(z)} \cdot dF^n(z) \ .$$

By (5.2) we have that

$$n\mathbf{E}R_n \le I_n \left[1 + o(1)\right] \qquad (n \to \infty) \ .$$

We therefore wish to show that

$$I_n \le \frac{\mathbf{E}X_{(n)}}{\mu} + o(\mathbf{E}X_{(n)}) \qquad (n \to \infty) \tag{5.11}$$

which will imply the result. Now

$$I_n \le \int_\delta^\infty \frac{z}{\mu(z)/F(z)} \cdot dF^n(z) \le \int_\delta^\infty \frac{z}{\mu(z)} \cdot dF^n(z) \tag{5.12}$$

Define

$$r(z) := \mu - \mu(z) \ .$$

Then from (5.12)

$$I_n \le \frac{1}{\mu} \int_\delta^\infty \frac{z}{1 - r(z)/\mu} \cdot dF^n(z) \tag{5.13}$$

Pick a δ so that $r(\delta)/\mu = \mu(\delta)/\mu = ½$. Then $F(\delta) < 1$, and

$$\forall z \ge \delta \quad \frac{r(z)}{\mu} \le \frac{1}{2}$$

and

$$\frac{1}{1 - r(z)/\mu} \le 1 + 2\frac{r(z)}{\mu} \ .$$

Using this δ in (5.13) yields

$$I_n \le \frac{1}{\mu} \int_\delta^\infty z \left[1 + 2\frac{r(z)}{\mu}\right] \cdot dF^n(z) = \frac{1}{\mu} \int_\delta^\infty z \cdot dF^n(z) + \frac{2}{\mu^2} \int_\delta^\infty z r(z) \cdot dF^n(z) \ .$$

Both integrals above are finite, since $zr(z) = o(z)$ as $z \to \infty$, and we have assumed that $\mathbf{E}X$ and therefore $\mathbf{E}X_{(n)}$ is finite.

We see from the above that

$$I_n \le \frac{\mathbf{E}X_{(n)}}{\mu} + \frac{2}{\mu} \int_0^\infty z \frac{r(z)}{\mu} \cdot dF^n(z) \tag{5.14}$$

Define $f(z) := zr(z)/\mu$. Observe that the second integral in (5.14) is just $(2/\mu) \cdot \mathbf{E}[fX_{(n)}]$ where f satisfies the hypotheses of Theorem (2.2.4). We have also assumed that F has unbounded support, so that $\mathbf{E}X_{(n)}$ is unbounded as $n \to \infty$. Therefore by Theorem (2.2.4), $\mathbf{E}[fX_{(n)}] = o(\mathbf{E}X_{(n)})$, and putting this into (5.14) leads to

$$I_n \le \frac{\mathbf{E}X_{(n)}}{\mu} + o(\mathbf{E}X_{(n)}) \qquad (n \to \infty) \ .$$

This is the required (5.11). □

6. Main Theorem

The main result may now be concluded:

THEOREM. Assume $X_1, \cdots, X_n$ are i.i.d. $=_d F$. Suppose either that

(i) $\bar{F} \in R_{-\alpha}$ for some $\alpha > 1$; or else

(ii) F has finite second moment.

Then

$$\mathbf{E}R_n = \frac{\mathbf{E}X_{(n)}}{\mathbf{E}S_n}\left[1 + o(1)\right] \qquad (n \to \infty) . \tag{6.1}$$

Proof: Suppose first that F has unbounded support. Then the lower bound in (6.1) follows from Theorem (4.1). Consider the upper bound. If hypothesis (i) applies, then $\bar{F} \in R_{-\alpha}$ for some $\alpha > 1$, and by Theorem (2.1.4), $\mathbf{E}[X_1^{1+\nu}] < \infty$ for any $\nu \in (0, \alpha - 1)$. Then Theorem (5.10) provides the upper bound. If hypotheses (ii) applies, then $\mathbf{E}[X_1^2] < \infty$ and again Theorem (5.10) supplies the upper bound.

It remains only to argue that the result holds when F has bounded support. Let $\omega < \infty$ be the upper support point of F. It is easy to see that

$$X_{(n)} \xrightarrow{as} \omega \qquad (n \to \infty) \tag{6.2}$$

Clearly $\mu < \infty$ in the case of bounded support, and by the strong law of large numbers

$$\frac{n}{S_n} \xrightarrow{as} \frac{1}{\mu} \qquad (n \to \infty) .$$

It follows that

$$nR_n \xrightarrow{as} \frac{\omega}{\mu} \qquad (n \to \infty) .$$

For all sufficiently large n, the sequence $\{nR_n\}$ is almost surely bounded by a constant—say $2\omega/\mu$. Under these uniformity conditions, it is well known [Chu74, Theorem 4.1.4] that convergence a.s. of a sequence implies convergence in moment of the same sequence, and so we conclude

$$\mathbf{E}[nR_n] \to \frac{\omega}{\mu} \qquad (n \to \infty) . \tag{6.3}$$

Since $X_{(n)} \le \omega$, the sequence in (6.2) is also uniformly bounded and hence we may similarly conclude from (6.2) that

$$\mathbf{E}[X_{(n)}] \to \omega \qquad (n \to \infty)$$

and hence that

$$\frac{\mathbf{E}[X_{(n)}]}{\mu} \to \frac{\omega}{\mu} \qquad (n \to \infty) . \tag{6.4}$$

Taken together, (6.3) and (6.4) show that

$$\mathbf{E}[nR_n] \sim \frac{\mathbf{E}X_{(n)}}{\mu} \qquad (n \to \infty) .$$

This completes the proof for the bounded support case, and with it, the proof of the Theorem. □

EXAMPLE. Suppose $X =_d F(t)$ obeys a Pareto law of shape p, where $F(t) = 1 - t^{-p}$ for $t \in [1, \infty)$. If $p > 1$, then $\mathbf{E}X^\nu < \infty$ for all $1 \le \nu < p$. A straightforward calculation [Dow90a] shows that

$$\mathbf{E}X_{(n)} \sim \Gamma(1 - 1/p) n^{1/p} \qquad (n \to \infty)$$

and it is easy to see that $\mathbf{E}X = 1/(p-1)$. By Theorem (6.1), since $\bar{F}$ is in R_{-p},

$$\forall p > 1 \quad \mathbf{E}R_n \sim \frac{\Gamma(1 - 1/p)}{p - 1} n^{-(1 - 1/p)} \qquad (n \to \infty) .$$

The main result is seen to be valid for all Pareto distributions for which a finite mean is well-defined. □

EXAMPLE. Refer to the uniform d.f. discussed in detail in the Example following Theorem (3.1). For X a $U(0, 1)$ variate, Theorem (6.1) provides information on the first term of the asymptotic expansion calculated there. Since $\mathbf{E}X_{(n)} = n^{-1}(n+1)$ and $\mathbf{E}S_n = n/2$ we have

$$\mathbf{E}R_n \sim \frac{2}{n} \qquad (n \to \infty) .$$

EXAMPLE. Let $X =_d F(t)$ where $F(t) = 1 - \exp[-x^{1/a}]$ is a Weibull d.f. with shape parameter $a > 0$. Standard methods in the asymptotic theory of extremes [Res87] yield $\mathbf{E}X_{(n)} = (\ln n)^a + O(\ln n)^{a-1}$. In this case, Theorem (6.1) yields, since $\mathbf{E}X = \Gamma(a+1)$

$$\mathbf{E}R_n \sim \frac{(\ln n)^a}{n\Gamma(a+1)} \qquad (n \to \infty) . \qquad \Box$$

7. References

[Amd67] G.M. AMDAHL, 1967. Validity of the single processor approach to achieving large scale computing capabilities, *AFIPS Conference Proceedings 30*, AFIPS Press, Reston, VA, pp. 483-485.

[Arn85] B.C. ARNOLD, 1985. *p*-norm bounds on the expectation of the maximum of a possibly dependent sample, *J. Multivariate Anal. 17*, 316-332.

[Arn89] B.C. ARNOLD and N. BALAKRISNAN, 1989. *Relations, Bounds and Approximations for Order Statistics*, Lecture Notes in Statistics No. 53, Springer Verlag, New York, NY.

[Aro60] D.Z. AROV and A.A. BOBROV, 1960. The extreme terms of a sample and their role in the sum of independent random variables, *Theory of Probability and Applications 5*, 377-96.

[Bar75] R.E. BARLOW and F. PROSCHAN, 1975. *Statistical Theory of Reliability and Life Testing: Probability Models*, Holt, Rinehart and Winston, New York, NY.

[Ber28] S. BERNSTEIN, 1928. Sur les fonctions absolument monotones, *Acta Mathematica 51*, 1-66.

[Bin81] N.H. BINGHAM and J.L. TEUGELS, 1981. Conditions implying domains of attraction, *Proceedings of the Sixth Conference on Probability Theory*, Brasov, Romania (Sep 10-15, 1979), Editura Academiei Republicii Socialiste Romania, Bucuresti, 1981, pp. 23-34.

[Bin87] N.H. BINGHAM, C.M. GOLDIE and J.L. TEUGELS, 1987. *Regular Variation*, volume 27 of *The Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, Cambridge, England.

[Bre65] L. BREIMAN, 1965. On some limit theorems similar to the arc-sine law, *Theory of Probability and Applications 10*, 323-31.

[Cho79] T.-L. CHOW and J.L. TEUGELS, 1979. The sum and the maximum of i.i.d. random variables, In *Asymptotic Statistics* (Proc. Second Prague Symposium, eds. P. Mandl and M. Huskova), North-Holland, Amsterdam, 81-92.

[Chu74] K.L. CHUNG, 1974. *A Course in Probability Theory, 2nd Ed.*, Academic Press, New York, NY.

[Cof76] E.G. COFFMAN, JR., 1976. *Computer and Job-Shop Scheduling Theory*, Wiley, New York, NY.

[Cof85] E.G. COFFMAN, JR. and E.G. GILBERT, 1985. On the expected relative performance of list scheduling, *Operations Research 33*, 548-561.

[Cof91] E.G. COFFMAN, JR. and G.S. LUEKER, 1991. *Probabilistic Analysis of Packing and Partitioning Algorithms*, Wiley, New York, NY.

[Dar52] D.A. DARLING, 1952. The role of the maximum term in the sum of independent random variables, *Trans. AMS 73*, 95-107.

[Dav70] H.A. DAVID, 1970. *Order Statistics*, Wiley, New York, NY.

[Dem83] M.A.H. DEMPSTER, M.L. FISHER, L. JANSEN, B.J. LAGEWEG, J.K. LENSTRA, and A.H.G. RINNOOY KAN, 1983. Analysis of heuristics for stochastic programming: results for hierarchical scheduling problems, *Mathematics of Operations Research 8*, 525-537.

[Dow90a] P.J. DOWNEY, 1990. Distribution-free bounds on the expectation of the maximum with scheduling applications, *Operations Research Letters 9*, 189-201.

[Dow90b] P.J. DOWNEY and R.S. MAIER, 1991. Stochastic orderings and extreme moment growth, Technical Report 90-9, Department of Computer Science, University of Arizona, Tucson, AZ.

[Dow91] P.J. DOWNEY and R.S. MAIER, 1991. Stochastic orderings and the mean growth of extremes, Technical Report 91-13, Department of Computer Science, University of Arizona, Tucson, AZ.

[Dow93] P.J. DOWNEY and R.S. MAIER, 1993. Orderings arising from expected extremes, with an application. *Stochastic Inequalities*, IMS Lecture Notes - Monograph Series, vol. 22, Institute of Mathematical Statistics, Hayward, CA, pp. 66-75.

[Dow94] P.J. DOWNEY, 1994. Bounds and approximations for overheads in the time to join parallel forks, *ORSA Journal on Computing 6*, in press.

[Fel66] W. FELLER, 1966. *An Introduction to Probability Theory and its Applications, Vol. II*, Wiley, New York, NY.

[Gar79] M.R. GAREY and D.S. JOHNSON, 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, CA.

[Gne43] B.V. GNEDENKO, 1943. Sur la distribution limité du terme d'une série aléatoire, *Ann. Math. 44*, 423-453.

[Gus88] J.L. GUSTAFSON, G.R. MONTRY, and R.E. BENNER, 1988. Development of parallel methods for a 1024-processor hypercube, *SIAM J. Sci. Stat. Comput. 9*, 609-638.

[Gra69] R.L GRAHAM, 1969. Bounds on multiprocessing timing anomalies, *SIAM J. of Appl. Math. 17*, 416-429.

[Gra76] R.L GRAHAM, 1976. Bounds on the performance of scheduling algorithms, in *Computer and Job-Shop Scheduling Theory*, E.G. Coffman, Jr. (ed.), Wiley, New York, NY.

[Kar30] J. KARAMATA, 1930. Sur un mode de croissance régulière des fonctions, *Mathematica (Cluj) 4*, 38-53.

[Kas84] Y. KASHARA, 1984. A note on sums and maxima of independent, identically distributed random variables, *Proc. Japan Acad. 60, A*, 353-6.

[Kum94] V. KUMAR and A. GUPTA, 1994. Analyzing scalability of parallel algorithms and architectures, *J. Parallel and Dist. Comp.*, Sep 1994, to appear.

[Law93]] E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN and D.B. SHMOYS, 1993. Sequencing and scheduling: algorithms and complexity, in *Handbooks in Operations Research and Management Science, Volume 4: Logistics of Production and Inventory*, S.C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin (eds.), North-Holland, Amsterdam, Netherlands, pp. 445-522.

[Lel86] W.E. LELAND and T.J. OTT, 1986. Load-balancing heuristics and process behavior, *Performance Evaluation Review 14*, 1(May 1986), 54-69.

[Len84] J.K. LENSTRA, A.H.G. RINNOOY KAN and L. STOUGIE, 1984. A framework for the probabilistic analysis of hierarchical planning systems, *Annals of Operations Research 1*, 23-42.

[Mal84] R.A. MALLER and S.I. RESNICK, 1984. Limiting behaviour of sums and the term of maximum modulus, *Proc. London Math. Soc. (3) 49*, 385-422.

[OBr80] G.L. O'BRIEN, 1980. A limit theorem for sample maxima and heavy branches in Galton-Watson trees, *J. Appl. Prob. 17*, 539-545.

[Olv74] F.W.J. OLVER, 1974. *Asymptotics and Special Functions*, Academic Press, New York, NY.

[Pic68] J. PICKANDS, 1968. Moment convergence of sample extremes, *Ann. Math. Statist. 39*, 881-889.

[Pru87] W.E. PRUITT, 1987. The contribution to the sum of the summand of maximum modulus, *Ann. Prob. 15*, 885-96.

[Res86] S.I. RESNICK, 1986. Point processes, regular variation and weak convergence, *Adv. Appl. Prob. 18*, 66-138.

[Res87] S.I. RESNICK, 1987. *Extreme Values, Regular Variation and Point Processes*, Springer-Verlag, New York, NY.

[Ros83] S.M. ROSS, 1983. *Stochastic Processes*, Wiley, New York, NY.

[Sto83] D. STOYAN, 1983. *Comparison Methods for Queues and Other Stochastic Models*, Wiley, New York, NY.

[Wid71] D.V. WIDDER, 1971. *An Introduction to Transform Theory*, Academic Press, New York, NY.

[Wid61] D.V. WIDDER, 1961. *Advanced Calculus, 2nd Ed.*, Prentice-Hall, Englewood Cliffs, NJ.