UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

ASYMPTOTIC BEHAVIOR OF THE EXPECTED VALUE OF ORDER STATISTICS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

MATÍAS ERNESTO ROMERO YÁÑEZ

PROFESOR GUÍA:
JOSÉ RAFAEL CORREA HAEUSSLER

MIEMBROS DE LA COMISIÓN:
DENIS ROLAND SAURÉ VALENZUELA
RAUL EDUARDO GOUET BAÑARES

Este trabajo ha sido parcialmente financiado por Proyecto de Anillo de Investigación ANID ACT210005, FONDECYT 1190043, el Centro de Modelamiento Matemático (ACE210010 and FB210005) y el Instituto Milenio para el Estudio de Imperfecciones del Mercado y Políticas Públicas (ICS13 002).

SANTIAGO DE CHILE
2022

RESUMEN DE LA MEMORIA PARA OPTAR
AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES
Y TÍTULO DE INGENIERO CIVIL MATEMÁTICO
POR: **MATÍAS ERNESTO ROMERO YÁÑEZ**
FECHA: 2022
PROF. GUÍA: JOSÉ RAFAEL CORREA HAEUSSLER

ASYMPTOTIC BEHAVIOR OF THE EXPECTED VALUE OF ORDER STATISTICS

En este trabajo estudiamos el comportamiento asintótico del valor esperado de los dos estadísticos de orden más grandes. Específicamente, buscamos describir la tasa de crecimiento más rápida posible de su valor esperado, cuando las variables aleatorias son independientes e idénticamente distribuidas. Para el máximo, demostramos de manera simple y directa que el crecimiento es sublineal y probamos que es imposible mejorar tal resultado. En contraste, en el caso del segundo estadístico de orden más grande derivamos una cota superior cuyo crecimiento es estrictamente más lento que el encontrado para el máximo, denotando un comportamiento asintótico considerablemente distinto en objetos aparentemente similares. Esta conclusión tiene consecuencias en contextos aplicados donde estos objetos modelan cantidades de interés, tales como scheduling estocástico, ingeniería de confiabilidad, gestión de riesgos, teoría de licitaciones, entre otras.

RESUMEN DE LA MEMORIA PARA OPTAR
AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES
Y TÍTULO DE INGENIERO CIVIL MATEMÁTICO
POR: **MATÍAS ERNESTO ROMERO YÁÑEZ**
FECHA: 2022
PROF. GUÍA: JOSÉ RAFAEL CORREA HAEUSSLER

ASYMPTOTIC BEHAVIOR OF THE EXPECTED VALUE OF ORDER STATISTICS

In this thesis we study the asymptotic behavior of the expected value of the two largest order statistics. Specifically, we seek to describe the fastest possible growth rate of their expected value, when the random variables are independent and identically distributed. For the maximum, we provide a simple and direct proof that the growth is sublinear, and we prove that it is impossible to improve such result. In contrast, in the case of the second largest order statistic we derive an upper bound with strictly slower growth rate than the one of the maximum, concluding considerably different asymptotic behavior in apparently similar objects. These conclusions have consequences in applied contexts where these objects model relevant quantities, such as stochastic scheduling, reliability, risk management, auction theory, among others.

Acknowledgements

En primer lugar quiero agradecer a mi familia. A mis padres Gonzalo y María Teresa por brindarme su máximo apoyo en todo mi proceso. A mi hermano Gonzalo y su familia Maca, Martina y Santi; mis hermanas Javi y Cami, y mi cuñado David, quienes fueron y siguen siendo un pilar fundamental en mi desarrollo personal y académico. Por último, a mi nani Lidia, a mi abuelita Carmen, y a todo el resto de mi familia que siempre ha estado presente y dispuestos a darme una mano con lo que necesite.

También agradezco A José Correa por ser un excelente profesor guía y gran inspiración sobre la vida académica y la investigación. También a los profesores Denis Sauré y Raul Gouet por integrar esta comisión y darme la oportunidad de enseñar en los cursos que más me interesaron durante la carrera. Asimismo agradezco a los demás profesores, funcionarios y funcionarias del DIM y MGO que fueron parte clave en mi formación académica y profesional. En especial a Linda, Fernanda, Olga, Karen, Eterin y Oscar, por su permanente disposición a ayudar en todo lo que sea necesario.

Por último quiero agradecer también a mis amistades. En especial al Rola, P., Barri y Rodrigo, con quienes compartimos la casa como lugar de vida, estudio y recreación. También a mi gente de Puerto Varas, con quienes en cada reencuentro vuelvo a sentir lo afortunado que fui de compartir mi infancia en tan hermoso lugar.

Contents

Chapter 1

Introduction

Years ago the study of order statistics was a curiosity [2]. Of course, it has always been introduced in any serious statistics course, but it was not until H.A. David published his first edition of *Order Statistics* [5] in 1969, that the theory, techniques, and applications of the subject became of growing recognition. Since that moment both the theory and applications of order statistics have greatly expanded. They have been widely used in many applied probability areas, such as stochastic scheduling, reliability, risk management, auction theory, among many others[10]. In particular, the asymptotic theory of extreme order statistics and of related statistics has been developed with increased emphasis. It does not only provide information for the asymptotic regimes, but also approximates probabilistic models for random quantities when the extremes govern the laws of interests (e.g. strength of materials, floods, droughts, etc.) [5, 9].

In this work, we study the asymptotic behavior of the expected value of the two largest order statistics. Specifically, we aim to describe their fastest growth rate possible in the setting of an arbitrarily large sample of independent and identically distributed random variables drawn from a common distribution. Our main results show that these two seemingly similar objects behave considerably different in the limit, which may be of interest in applied probability contexts where the expectation of these order statistics represent relevant quantities.

1.1 Order Statistics

Suppose that $X_1, \ldots, X_n$ are n random variables. The k-th order statistic, denoted by $X_{k:n}$, is the k-th smallest one of them, that is, they can be arranged in order of magnitude as follows:

$$X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}.$$

Although this definition does not require the random variables to be independent nor identically distributed, these assumptions are usually considered, since most of the classical results dealing with order statistics were originally derived in that setting [2]. For instance, in such setting we can easily derive the cumulative distribution function of $X_{k:n}$. Let F be the common parent distribution for $X_1, \dots, X_n$, and consider the event $E_{\mathrm{i},n}(x)$ that exactly i of the n i.i.d random variables are less than or equal to x. Then,

$$\begin{aligned} F_{k:n}(x) &:= \mathbb{P}(X_{k:n} \leq x) \\ &= \sum_{\mathrm{i}=k}^{n} \mathbb{P}(E_{\mathrm{i},n}(x)) \\ &= \sum_{\mathrm{i}=k}^{n} \binom{n}{\mathrm{i}} F(x)^{\mathrm{i}}(1-F(x))^{n-\mathrm{i}}. \end{aligned}$$

We work in this i.i.d. setting throughout this thesis, and also assuming that the random variables are non-negative. This last assumption is not restrictive, since the bound $X_{k:n} \leq |X_{k:n}|$ allow us to replicate all the results, provided that $\mathbb{E}(|X_1|) < \infty$. Also, the non-negativity assumption is usually considered in most of the the applied contexts that concern this thesis work.

1.1.1 Asymptotic Analysis

The asymptotic theory of order statistics is concerned with the properties of $X_{k:n}$ as n tends to infinity. In general, it aims to replace or approximate complicated situations in probabilistic models by a comparatively simple asymptotic model. A lot can be said about the asymptotic distributions, theory of extremes and even extremal processes, which have been developed at length by Galambos in [9]. Most of the results depend on $p = \lim_{n\to\infty} k/n \in [0,1]$ and are fundamentally different among three major categories [5]:

1. *Central or quantile case:* If $p \in (0,1)$.
2. *Extreme case:* If $p \in \{0,1\}$, for fixed k. $X_{k:n}$ and $X_{n-k+1:n}$ are usually called the k-th lower and upper extremes, respectively.
3. *Intermediate case:* If $p \in \{0,1\}$, with k being a function of n.

The problem we address is to determine the fastest growth rate of the expectation of the first two upper extremes in the case of finite expectation distributions, that is, for $k \in \{1,2\}$,

$$\max_{\substack{F \text{ with} \\ \text{finite expectation}}} \mathbb{E}(X_{n-k+1:n}).$$

To describe the limiting behavior, we use standard notation for asymptotic analysis. Given two functions $f, g : \mathbb{R} \to \mathbb{R}_+$,

- $f(x) = O(g(x))$ if there exists $C > 0, x_0 > 0$ such that $f(x) \leq Cg(x)$ for all $x \geq x_0$.
- $f(x) = \Omega(g(x))$ if there exists $C > 0, x_0 > 0$ such that $f(x) \geq Cg(x)$ for all $x \geq x_0$.
- $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.
- $f(x) \sim g(x)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

1.1.2 Some Applications

The theoretical and practical interest in some order statistics is quite natural. This is the case of the extremes, $X_{1:n}$ (minimum) and $X_{n:n}$ (maximum), or the median, denoted as $X_{\frac{n+1}{2}:n}$ if n is odd or $(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n})/2$ if n is even, which are common topics in any statistics course. However, the scope of the applications of order statistics in data-analytic models goes beyond that.

The expected value of the maximum is a canonical benchmark in many applied probability subjects. For instance, in the context of stochastic scheduling, it represents the expected completion time to run n tasks in n parallel processors, each with a random processing time. In contrast, the expected completion time of these tasks executed sequentially is the sum of each processor's expected time, which in the i.i.d. case results in a linear function of the number of processors n. In this setting, to study the fraction of the time that can be saved by parallel computation as the number of processors increases, it is important to know how fast may the expected value of the maximum grow as the number of samples goes to infinity [6]. Similarly, in the context of prophet inequalities this question is relevant as well. In the general statement, a gambler faces an ordered sequence of non-negative independent random variables and must decide online when to stop in order to maximize expected reward. In the i.i.d. setting, the gambler obtains at least a $1 - 1/\mathrm{e}$ fraction of the reward of a prophet who knows all the values and can choose the largest one [4]. This lower bound can be achieved even by a specific class of strategies called *fixed threshold algorithms* (see [7] for details on FTA). Moreover, this fraction is best possible in this setting, in the sense that given a sequence of length n, there is a distribution dependent on n for which no FTA can achieve an approximation factor better than $1 - 1/\mathrm{e}$. However, this may not hold true if we consider an arbitrarily large sequence and a distribution that does not depend on n. To study this question and look for a worst case, the fastest growth rate possible of the expected maximum is a relevant element.

Another interesting example arises in the study of *reliability* of systems. In that context, a k-out-of-n system is a system consisting of n components and normally operating if and only if at least k of the n components work. Therefore, the lifetime of this system can be represented by the k-th upper extreme order statistic $X_{n-k+1:n}$. As the very popular fault tolerant structure, the k-out-of-n system has been widely applied in industrial engineering and electrical systems [10]. Particularly, the n-out-of-n and the 1-out-of-n systems correspond to the *series* and *parallel* systems, respectively. Again, in this context, the growth rate of the expectation of order statistics is key to understand the expected lifetime of these systems as the number of components increases.

1.2 Outline and Contributions of Our Work

Motivated by the relevant applications, we study the fastest growth rate of the expected value of the first two upper extremes, in the independent and identically distributed setting. In Chapter 2, we elaborate a simple alternative proof of the known fact that the expectation of the maximum is sublinear, in the sense that it grows slower than any linear function.

The approach is rather elementary and elegant, requiring only a few lines, in contrast to the original rather long proof. Additionally, in Section 2.3 we prove an impossibility result that allow us to conclude that this statement cannot be improved.

In Chapter 3, we continue to study the second upper extreme order statistic. The reduction in Section 3.1 allows us to identify the differences of the asymptotic properties of the second upper extreme problem. In fact, in Section 3.2 we specify the properties that are used in Section 3.3 to conclude that the asymptotic behavior of the expected value of the second upper extreme is fundamentally different. Specifically, its fastest growth rate is faster than the one of the maximum, meaning that for some distributions, the fraction that it gets over the expected maximum goes to zero as the number of samples increases. Finally, in Section 3.4 we adapt the methodology from the previous chapter to obtain a lower bound for the fastest growth rate possible and leave the improvements of this bound as an open question.

Chapter 2

Maximum or First Upper Extreme

In this chapter we study the asymptotic behavior of the expectation of the maximum order statistic $X_{n:n}$. We provide an elementary proof for an $o(n)$ upper bound for $\mathbb{E}(X_{n:n})$, which only uses the dominated convergence theorem and a basic calculus result. As a corollary, we also obtain that if moments of high order are finite, the fastest growth rate possible of $\mathbb{E}(X_{n:n})$ gets slower. Finally, we construct a distribution with finite expectation that allow us to conclude that the $o(n)$ bound is indeed best possible.

2.1 Preliminaries

For fixed $n > 1$, suppose $X_1, \ldots, X_n$ are independently drawn from a fixed distribution F with finite expectation, and denote $X_{n:n} = \max_{\mathrm{i}=1,\ldots,n} X_{\mathrm{i}}$. Note that since the maximum is upper bounded by the sum, we have that $\mathbb{E}(X_{n:n}) \leq n\mathbb{E}(X_1)$, so $\mathbb{E}(X_{n:n}) = O(n)$. Similarly if $\mathbb{E}(X_1^p) < \infty$, it is easy to derive that $\mathbb{E}(X_{n:n}) = O(\sqrt[p]{n})$ using Jensen's inequality. This means that for distributions with finite absolute moments of high order, $\mathbb{E}(X_{n:n})$ grows very slowly in n. Moreover, for some particular distributions such as the exponential, it is easy to see that $E(X_{n:n}) = O(\log n)$, and for the Gaussian distribution $E(X_{n:n}) = O(\sqrt{\log n})$.

In general, when the distribution from where $X_1, \ldots, X_n$ are drawn can depend on n, explicit upper bounds obtained by e.g. Arnold [1] and Downey [6] can be achieved by a suitable extremal distribution. However, when F is fixed and does not depend on n a much stronger and general bound can be obtained. Indeed, Downey [6] established that $\mathbb{E}(X_{n:n}) = o(n)$.[1]

To illustrate the difference between this statement and the previous $O(n)$ upper bound, recall the problem presented in the introduction about the fraction of time that is saved by parallel computation. Suppose $X_1, \ldots, X_n$ are i.i.d. processing times of the n processors. For fixed n, the fraction of the expected time for sequential processing that takes the parallel computation, that is $\frac{\mathbb{E}(X_{n:n})}{n\mathbb{E}(X_1)}$ can be constant in n if the distribution of the processing times depends on n. However, there is no reason to think that the distribution of each processing

[1]More generally Downey establishes that if $\mathbb{E}(X_1^p) < \infty$ then $\mathbb{E}(X_{n:n}) = o(\sqrt[p]{n})$.

time depends on the number of processors. Therefore, as F should be independent of n, the much stronger result $\mathbb{E}(X_{n:n}) = o(n)$ implies that the fraction becomes arbitrarily small as the number of processors increases.

To establish that $\mathbb{E}(X_{n:n}) = o(n)$, Downey follows several steps. He first studies the sequence $(X_{n:n}/\sqrt[p]{n})_n$, and uses a result by Freedman [8] to establish its convergence in probability, that is, for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(X_{n:n} > \sqrt[p]{n}\varepsilon) = 0.$$

Then he turns to prove that the sequence also converges in L^p. To this end, he shows that for $p = 1$ the sequence is uniformly integrable and thus, by Vitali convergence theorem (see e.g. [3, Theorem 4.5.4]), obtains L^1-convergence.

For general $p \geq 1$, and under the assumption $\mathbb{E}(X^p) < \infty$, Downey resorts to Hölder inequality to reduce to the $p = 1$ case and concludes that $\mathbb{E}(X_{n:n}) = o(\sqrt[p]{n})$.

Our proof for the $o(n)$ upper bound relies only on the fact that the arithmetic mean of a convergent sequence converges to the same limit. This preliminary result can be found in [11] as a consequence of the Stolz–Cesàro theorem. Here we provide a direct proof of it.

Lemma 2.1 *Let $(x_n)_n$ be a sequence of real numbers which converges to $l \in \mathbb{R}$. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k = l$$

Proof. Note first that it is enough to consider $l = 0$, for otherwise we may take $y_k = x_k - l$ which converges to 0 and

$$\frac{1}{n}\sum_{k=1}^{n} y_k = \left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) - l.$$

Let $\varepsilon > 0$ and take $N \in \mathbb{N}$ such that $|x_k| < \varepsilon$ for all $k > N$. Then for $n > N$

$$\left|\frac{1}{n}\sum_{k=1}^{n} x_k\right| < \frac{N}{n}\max_{k=1,\dots,N}|x_k| + \frac{(n-N)}{n}\varepsilon,$$

which becomes less than 2ε for all sufficiently large n. □

2.2 Sublinearity of the Expected Maximum

In this section, we present our proof for the $o(n)$ upper bound and use Jensen's inequality to state it in Downey's general form.

Theorem 2.2 *Let $X_1, \dots, X_n$ be independent random variables drawn from a common distribution F. Suppose $\mathbb{E}(X_1) < \infty$, then*

$$\lim_{n\to\infty} \frac{\mathbb{E}(X_{n:n})}{n} = 0.$$

Proof. Since $X_{n:n}$ is a non-negative random variable with distribution F^n, its expectation can be written as

$$\mathbb{E}(X_{n:n}) = \int_0^\infty 1 - F^n(x)\mathrm{d}x = \int_0^\infty (1 - F(x)) \sum_{k=1}^{n-1} F^k(x)\mathrm{d}x.$$

The linearity of the integral implies that

$$\frac{\mathbb{E}(X_{n:n})}{n} = \frac{1}{n} \sum_{k=1}^{n-1} \int_0^\infty F^k(x)(1 - F(x))\mathrm{d}x.$$

To conclude the proof, recall that by Lemma 2.1, it is enough to argue that

$$\lim_{n\to\infty} \int_0^\infty F^n(x)(1 - F(x))\mathrm{d}x = 0.$$

This follows by the dominated convergence theorem since the sequence $(F^n(1-F))_n$ converges pointwise to 0 and it is dominated by the integrable function $1 - F$.[2] □

Note that by Vitali convergence theorem, the L^1-convergence of the sequence $(X_{n:n}/n)_n$, is equivalent to its convergence in probability, and also to its uniform integrability. Therefore, the convergence in expectation we just showed also implies the convergence in probability and the uniform integrability shown by Downey. Furthermore, using Jensen's inequality for a convex function h we get

$$h(\mathbb{E}(X_{n:n})) \leq \mathbb{E}(h(X_{n:n})) \leq \mathbb{E}\left(\max_{\mathrm{i}=1,\ldots,n} h(X_\mathrm{i})\right).$$

For instance, if $h(x) = x^p$ for any $p \geq 1$, then from Theorem 2.2 we get that if $\mathbb{E}(X_1^p) < \infty$, then $\mathbb{E}(X_{n:n})^p \leq \mathbb{E}(\max_{\mathrm{i}=1,\ldots,n} X_\mathrm{i}^p) = o(n)$. Thus we immediately obtain the following more general result as a corollary.

Theorem 2.3 *For any convex function h, if $\mathbb{E}(h(X_1)) < \infty$, then $h(\mathbb{E}(X_{n:n})) = o(n)$. In particular, for all $p \geq 1$, if $\mathbb{E}(X_1^p) < \infty$, then $\mathbb{E}(X_{n:n}) = o(\sqrt[p]{n})$.*

2.3 Impossibility Result

Downey [6] stated that the $o(n)$ upper bound is best possible in the following sense. He showed that for all $\varepsilon > 0$, there exists a distribution F, such that $\mathbb{E}(X_{n:n}) = \Omega(n^{1-\varepsilon})$. To illustrate this, consider $X_1, \ldots, X_n$ drawn from a Pareto distribution with scale 1 and shape $\alpha = (1-\varepsilon)^{-1} > 1$. Then, $\mathbb{E}(X_1) = 1 - 1/(\alpha - 1) < \infty$ and using the fact that for $x \leq n^{1/\alpha}$ $(1 - 1/x^\alpha)^n \leq (1 - 1/n)^n \leq 1/\mathrm{e}$ we get

$$\mathbb{E}(X_{n:n}) = \int_0^\infty 1 - \left(1 - \frac{1}{x^\alpha}\right)^n \mathrm{d}x \geq \int_0^{n^{1/\alpha}} 1 - \frac{1}{\mathrm{e}}\mathrm{d}x = \left(1 - \frac{1}{\mathrm{e}}\right) n^{1-\varepsilon}.$$

[2] Note that the sequence actually decreases to 0, so monotone convergence can also be invoked.

However, this does not rule out the possibility of having a result stronger than that in Theorem 2.2, such as $\mathbb{E}(X_{n:n}) = O(n/\log(n))$. In this section, we prove that it is impossible to improve the bound from Theorem 2.2.

Theorem 2.4 *For any function g with sublinear growth, namely such that $g(n) = o(n)$, there is a finite expectation distribution F such that if $X_1, \dots, X_n$ are independently drawn from F, then*

$$\liminf_{n\to\infty} \frac{\mathbb{E}(X_{n:n})}{g(n)} > 0.$$

Proof. We establish the statement by constructing a distribution F such that for all sufficiently large n.

$$\mathbb{E}(X_{n:n}) \geq g(n).$$

Without loss of generality, we assume that g is non-decreasing and has non-increasing increments, that is, $g(k+1) - g(k) \leq g(k) - g(k-1) \geq 0$ for all $k \geq 1$. For otherwise we may take

$$\tilde{g}(n) = g(0) + \sum_{k=1}^{n} \max_{m\geq k}(g(m) - g(m-1))$$

which satisfies such properties, along with $\tilde{g}(n) \geq g(n)$ and $\tilde{g}(n) = o(n)$. Indeed, by Proposition 2.1 we get

$$\lim_{n\to\infty} \frac{\tilde{g}(n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \max_{m\geq k}(g(m) - g(m-1)) = \limsup_{k\to\infty} g(k) - g(k-1) = 0,$$

and it is enough to show $\mathbb{E}(X_{n:n}) \geq \tilde{g}(n)$. Also, as we only need to show the inequality for sufficiently large n, we may assume that $g(0) \geq 0$.

Therefore we construct a distribution F of the form

$$F(x) = \sum_{k\geq 0} \left(1 - \frac{1}{k}\right) \mathbf{1}_{I_k}(x),$$

for some disjoint intervals $I_k \subseteq \mathbb{R}$ with length

$$\delta_k = \left(\frac{g(k)}{k} - \frac{g(k+1)}{k+1}\right) k$$

To establish that $\delta_k \geq 0$ we need to show that $(k+1)g(k) - kg(k+1) \geq 0$. This can be seen by induction over $k \geq 0$: the base case is $g(0) \geq 0$, and the non-increasing increments imply $g(k+1) \leq 2g(k) - g(k-1)$, thus

$$(k+1)g(k) - kg(k+1) \geq (k+1)g(k) - k(2g(k) - g(k-1)) = kg(k-1) - (k-1)g(k)$$

which is non-negative due to the inductive hypothesis.

With the previous choice of δ_k we immediately get that F has finite expectation. Indeed,

$$\int_0^\infty (1 - F(x))\mathrm{d}x = \sum_{k\geq 0} \frac{\delta_k}{k} = \sum_{k\geq 1} \frac{g(k)}{k} - \frac{g(k+1)}{k+1} = g(1) < \infty.$$

On the other hand, if $Y_1, \dots, Y_n$ are independent random variables drawn from F, we have that

$$\mathbb{E}(Y_{(n)}) = \int_0^\infty (1 - F^n(x))\mathrm{d}x = \sum_{k\geq 0} \left(1 - \left(1 - \frac{1}{k}\right)^n\right) \delta_k.$$

To wrap up the proof we lower bound the latter expression. First recall that $(1-1/x)^x$ grows to e^{-1} as $x \to \infty$. Also, from the strict convexity of the exponential function, we have that if $x \in (0,1)$, then $\exp(-x) < 1 - (1-\mathrm{e}^{-1})x$. Thus, for all $k \geq n$ we obtain

$$\left(1 - \frac{1}{k}\right)^n = \left(\left(1 - \frac{1}{k}\right)^k\right)^{n/k} \leq \exp(-n/k) < 1 - (1-\mathrm{e}^{-1})\frac{n}{k}.$$

Putting all together we derive the lower bound

$$\mathbb{E}(Y_{(n)}) > (1-\mathrm{e}^{-1})n \sum_{k\geq n} \frac{\delta_k}{k} = (1-\mathrm{e}^{-1})n \sum_{k\geq n} \frac{g(k)}{k} - \frac{g(k+1)}{k+1} = (1-\mathrm{e}^{-1})g(n),$$

which proves the statement by taking $X_\mathrm{i} = Y_\mathrm{i}/(1-\mathrm{e}^{-1})$. □

Chapter 3

Second Upper Extreme

In this chapter study the behavior of the second upper extreme order statistic. We provide an upper bound for $\mathbb{E}(X_{n-1:n})$ with a different asymptotic behavior from that of the maximum. To accomplish that, we reduce the problem to consider only distributions of the form used for the impossibility result in the previous chapter and study the differences in their structure. Additionally, we derive a lower bound for the fastest growth rate possible through the adaptation of the methodology from the previous chapter.

3.1 Reduction

Inspired by the distribution constructed for the lower bound of the previous chapter, we define the following family of distributions

Definition 3.1 *We denote by $\mathcal{F}$ the family of distributions F such that*

$$F(x) = \sum_{k \geq 1} \left(1 - \frac{1}{k}\right) \mathbf{1}_{I_k}(x),$$

for some disjoint intervals $I_k \subseteq \mathbb{R}$ with length $\delta_k > 0$.

We continue to prove that it is enough to consider only this family in order to find an upper bound for the expected value of $X_{n-1:n}$.

Proposition 3.2 *Let F be any distribution with finite expectation and let $X_1, \ldots, X_n$ be drawn from F. Then, there is a distribution $\tilde{F} \in \mathcal{F}$ with finite expectation such that if $Y_1, \ldots, Y_n$ are drawn from $\tilde{F}$,*

$$\mathbb{E}(X_{n-1:n}) \leq \mathbb{E}(Y_{n-1:n}).$$

Proof. For $k \geq 1$ define $x_k = \inf\{x : F(x) \geq 1 - 1/k\}$ and

$$\tilde{F}(x) = \sum_{k \geq 2} \left(1 - \frac{1}{k}\right) \mathbf{1}_{(x_k, x_{k+1}]}(x).$$

It is clear that $\tilde{F} \leq F$ and therefore $\mathbb{E}(X_{n-1:n}) \leq \mathbb{E}(Y_{n-1:n})$. In addition, we have that

$$\begin{aligned}
\int_0^\infty F(x) - \tilde{F}(x)\mathrm{d}x &= \sum_{k\geq 2} \int_{x_k}^{x_{k+1}} F(x) - \left(1 - \frac{1}{k}\right)\mathrm{d}x \\
&\leq \sum_{k\geq 2} \left(\frac{1}{k} - \frac{1}{k+1}\right)(x_{k+1} - x_k) \\
&= \sum_{k\geq 2} \frac{1}{k(k+1)}(x_{k+1} - x_k) \\
&\leq \sum_{k\geq 2} \frac{1}{(k+1)}(x_{k+1} - x_k) \\
&= \sum_{k\geq 2} \int_{x_k}^{x_{k+1}} \frac{1}{(k+1)}\mathrm{d}x \\
&\leq \int_0^\infty 1 - F(x)\mathrm{d}x = \mathbb{E}(X_1).
\end{aligned}$$

where we used that $F(x) \leq 1 - \frac{1}{k+1}$ for all $x \in (x_k, x_{k+1}]$. Thus,

$$\int_0^\infty 1 - \tilde{F}(x)\mathrm{d}x = \int_0^\infty 1 - F(x)\mathrm{d}x + \int_0^\infty F(x) - \tilde{F}(x)\mathrm{d}x \leq 2\mathbb{E}(X_1) < \infty,$$

and $\tilde{F}$ has finite expectation. It is easy to see that the previous statement applies for any other order statistic, since the inequality $\mathbb{E}(X_{k:n}) \leq \mathbb{E}(Y_{k:n})$ holds for any k. □

3.2 Structure, Differences and Asymptotic Properties

Recall from the previous chapter that if $F \in \mathcal{F}$ and $X_1, \ldots, X_n$ are drawn from F then

$$\mathbb{E}(X_{n:n}) = \sum_{k\geq 1} \left(1 - \left(1 - \frac{1}{k}\right)^n\right)\delta_k = \sum_{k\geq 1} k\left(1 - \left(1 - \frac{1}{k}\right)^n\right)\frac{\delta_k}{k}.$$

with $\sum \delta_k/k < \infty$. Similarly, using the fact that the distribution of $X_{n-1:n}$ is $F^n + nF^{n-1}(1 - F)$, we get that

$$\begin{aligned}
\mathbb{E}(X_{n-1:n}) &= \int_0^\infty 1 - F^n(x) - nF^{n-1}(x)(1 - F(x))\mathrm{d}x \\
&= \sum_{k\geq 1} \left(1 - \left(1 - \frac{1}{k}\right)^n - n\left(1 - \frac{1}{k}\right)^{n-1}\frac{1}{k}\right)\delta_k \\
&= \sum_{k\geq 1} \left(k - \left(1 - \frac{1}{k}\right)^{n-1}(k - 1 + n)\right)\frac{\delta_k}{k}.
\end{aligned}$$

This suggests that the key differences of their asymptotic behavior are represented by the properties of the functions $\varphi_n(x) = x(1-(1-1/x)^n)$ and $\phi_n(x) = x-(1-1/x)^{n-1}(x+n-1)$

(see Figure 3.1). The following results summarizes relevant properties that will be useful in the proof of the main theorem.

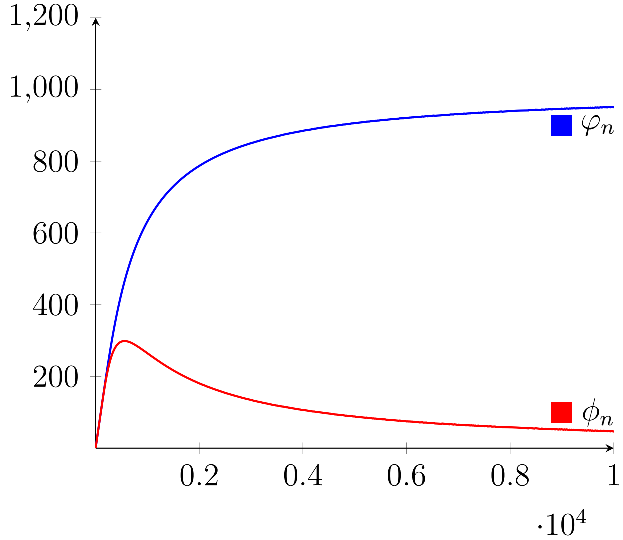

Figure 3.1: Plot of ϕ_n and φ_n for $n = 1000$.

Proposition 3.3 *Let* $\varphi_n(x) = x(1-(1-1/x)^n)$ *and* $\phi_n(x) = x - (1-1/x)^{n-1}(x+n-1)$ *defined for* $x \in [1, \infty)$ *and* $n \geq 2$.

(i) For all $n \geq 2$, $\varphi_n(x)$ *converges to* n *as* x *grows to infinity.*

(ii) For all $n \geq 2$, $\phi_n(x)$ *converges to* 0 *as* x *grows to infinity.*

Proof. By the Binomial Theorem,

$$\left(1-\frac{1}{x}\right)^n = \sum_{k=0}^{n} \binom{n}{k} (-x)^{-k} = 1 - \frac{n}{x} + o(1/x).$$

Thus, $\varphi_n(x) = n - xo(1/x)$ tends to n as x increases. Similarly,

$$(1-1/x)^{n-1}(x+n-1) = (x+n-1)\left(1-\frac{n-1}{x}+o(1/x)\right) = x - \frac{(n-1)^2}{x} + (x+n-1)o(1/x).$$

Therefore $\phi_n(x) = (n-1)^2/x - (n-1)o(1/x) - xo(1/x)$ converges to zero as x tends to infinity. □

Proposition 3.3 describes the main difference from the problem of the previous chapter. The impossibility result for the maximum relies on the fact that $\varphi_n(x) \geq (1 - \mathrm{e}^{-1})n$ for all $x > n$. Not only does this not hold true anymore, but the convergence to zero prevent us from deriving any useful lower bound up to infinity. This behavior allows us to improve the $o(n)$ upper bound by dividing the domain of ϕ_n in intervals where it can be bounded by some explicit function of n.

Proposition 3.4 *For $n \geq 3$, $\phi_n(x)$ is non-increasing in $[n-1, \infty)$.*

Proof. Note that

$$\phi_n(x) = x - \left(\frac{x-1}{x}\right)^{n-1}(x+n-1) = \frac{(x-1)^n}{x^{n-1}}\left(\left(\frac{x}{x-1}\right)^n - 1 - \frac{n}{x-1}\right),$$

and by the Binomial Theorem

$$\left(\frac{x}{x-1}\right)^n = \left(1 + \frac{1}{x-1}\right)^n = 1 + \frac{n}{x-1} + \sum_{k=2}^{n}\binom{n}{k}\left(\frac{1}{x-1}\right)^k.$$

Hence,

$$\phi_n(x) = \sum_{k=2}^{n}\binom{n}{k}\frac{(x-1)^{n-k}}{x^{n-1}}.$$

Now, since for each $k = 2, \dots, n$, the function $x \mapsto \frac{(x-1)^{n-k}}{x^{n-1}}$ decreases after its critical point $x = \frac{n-1}{k-1}$, we have that $\phi_n(x)$ is decreasing for $x \geq n-1$. □

Proposition 3.5 *For $n \geq 20$, $\phi_n(x)$ is non-decreasing in $[1, n/20]$.*

Proof. To prove this statement, note that the derivative of $\phi_n(x)$ can be written as

$$\phi_n'(x) = 1 - \left(1 - \frac{1}{x}\right)^{n-2}\frac{x^2 + (n-2)x + (n-1)^2}{x^2}.$$

We prove that $\phi_n'(x) \geq 0$ in $[1, n/20]$ by showing that the second term, $f_n(x) = (1 - 1/x)^{n-2}(x^2 + (n-2)x + (n-1)^2)/x^2$, is non-decreasing in that interval, and $f_n(n/20) < 1$. Define

$$h_n(x) = \log(f_n(x)) = (n-2)\log\left(1 - \frac{1}{x}\right) + \log(x^2 + (n-2)x + (n-1)^2) - 2\log(x),$$

then the derivative is

$$h_n'(x) = \frac{n-2}{x(x-1)} + \frac{2x + (n-2)}{x^2 + (n-2)\,x + (n-1)^2} - \frac{2}{x} = \frac{-n(n-1)x + n(n-1)^2}{x(x-1)(x^2 + (n-2)x + (n-1)^2)} \geq 0,$$

for $x \leq n-1$, since the denominator is non-negative and the numerator is a linear function with negative slope and solution at $x = n-1$. Hence, $h_n(x)$ and consequently $f_n(x)$ is non-decreasing in $[1, n/20]$. Finally,

$$f_n(n/20) \leq 421\left(1 - \frac{20}{n}\right)^{n-2} \leq 421(\mathrm{e}^{-20})^{1-2/n} \leq 421\mathrm{e}^{-20/3} \approx 0.53 < 1.$$

□

3.3 Main Theorem

We present the main result as follows:

Theorem 3.6 *Let $X_1, \dots, X_n$ be independent random variables drawn from a common distribution F. Suppose $\mathbb{E}(X_1) < \infty$, then*

$$\liminf_{n\to\infty} \frac{\mathbb{E}(X_{n-1:n})}{\frac{n}{\log(\log(n))}} = 0.$$

The theorem states that if such limit exists, it must be zero, in which case $\mathbb{E}(X_{n-1:n})$ is $o\left(n/\log(\log(n))\right)$. This behavior is fundamentally different from the one of the maximum, since the latter had no chance to be improved from the $o(n)$ bound.

To illustrate the difference, recall from Theorem 2.4 that for any function with sublinear growth, in particular $g(n) = n/\log(\log(n))$, we can find a distribution F such that $\mathbb{E}(X_{n:n}) \geq g(n)$ when $X_1, \dots, X_n$ are draw from F. If the limit in Theorem 3.6 exists for some distribution G, then if $Y_1, \dots, Y_n$ are drawn from G we get that

$$\lim_{n\to\infty} \frac{\mathbb{E}(Y_{n-1:n})}{\mathbb{E}(X_{n:n})} = \lim_{n\to\infty} \frac{\mathbb{E}(Y_{n-1:n})}{g(n)} \frac{g(n)}{\mathbb{E}(X_{n:n})} = 0,$$

since the second term is bounded. This argument can be replicated with any function g with sublinear growth and such that $g(n) = \Omega\left(n/\log(\log(n))\right)$. This conclusion may be useful in any applied probability context that benefits from a comparison between the two largest order statistics. For instance, in the context of reliability, if the number of components is arbitrarily large, the expected lifetime of a 2-out-of-n system could be arbitrarily shorter than that of a parallel system.

Proof of Theorem 3.6. Given the reduction of Proposition 3.2, it is enough to work with $X_1, \dots, X_n$ drawn independently from $F \in \mathcal{F}$, such that

$$\mathbb{E}(X_{n-1:n}) = \sum_{k\geq 1} \phi_n(k) \frac{\delta_k}{k},$$

for an arbitrary sequence $(\delta_k)_k$ such that $\mathbb{E}(X_1) = \sum \delta_k/k < \infty$.

Let $0 < \varepsilon < 1$ and consider $N_1(n) = n^{1-\varepsilon}$, $N_2(n) = n^{1+\varepsilon}$. Recall from Propositions 3.4 and 3.5 that for $n \geq 3$, $\phi_n(k)$ is non-decreasing if $k \leq \lceil N_1(n) \rceil \leq n/20$ and non-increasing if $x \geq \lfloor N_2(n) \rfloor \geq n-1$. Therefore

$$\sum_{k\geq 1} \phi_n(k) \frac{\delta_k}{k} \leq \phi_n(\lceil N_1(n) \rceil) \sum_{k=1}^{\lceil N_1(n) \rceil} \frac{\delta_k}{k} + \sum_{k=\lceil N_1(n) \rceil}^{\lfloor N_2(n) \rfloor} \phi_n(k) \frac{\delta_k}{k} + \phi_n(\lfloor N_2(n) \rfloor) \sum_{k\geq \lfloor N_2(n) \rfloor} \frac{\delta_k}{k}.$$

The sums in the first and third term can be bounded by $\mathbb{E}(X_1)$. For the middle term, recall that $\phi_n(x) \leq \varphi_n(x) \leq n$ for all x. Thus,

$$\mathbb{E}(X_{n-1:n}) \leq \max\left\{\phi_n(\lceil N_1(n) \rceil), \phi_n(\lfloor N_2(n) \rfloor)\right\} \mathbb{E}(X_1) + n \sum_{k=\lceil N_1(n) \rceil}^{\lfloor N_2(n) \rfloor} \frac{\delta_k}{k}. \tag{3.1}$$

Define $b_n = \int_n^\infty 1 - F$ and note that the sum in (3.1) is bounded by $\Delta_n = b_{N_1(n)} - b_{N_2(n)}$, which converges to 0 as n goes to infinity. In order to study its rate of convergence, we use the fact that for any non-negative sequence $(a_n)_n \subseteq \mathbb{R}$ such that $\sum a_n < \infty$, we have $\liminf_{n\to\infty} na_n = 0$.[1] Specifically, we find a subsequence $(\Delta_{n(k)})_k$ such that $\sum \Delta_{n(k)} < \infty$.

Consider $n(k) = 2^{\left(\frac{1+\varepsilon}{1-\varepsilon}\right)^k}$, then $N_1(n(k+1)) = N_2(n(k))$ and

$$\sum_{k\geq 1} \Delta_{n(k)} = \sum_{k\geq 1} b_{N_1(n(k))} - b_{N_1(n(k+1))} = b_{N_1(n(1))} - \lim_{k\to\infty} b_{N_1(n(k))} \leq \mathbb{E}(X_1) < \infty.$$

Hence,

$$\liminf_{k\to\infty} k\Delta_{n(k)} = 0,$$

and since $(k\Delta_{n(k)})_k$ is a subsequence of $(\log(\log(n))\Delta_n)$ we get

$$\liminf_{n\to\infty} \log(\log(n))\Delta_n = 0.$$

To wrap up the proof, recall that $\phi_n(x) \leq x$, so $\phi_n(N_1(n)) \leq N_1(n) = n^{1-\varepsilon}$. In addition, since $(1 - 1/n^{1+\varepsilon})^{n-1} \sim \mathrm{e}^{-1/n^\varepsilon} \sim 1 - 1/n^\varepsilon$, we have that

$$\phi_n(N_2(n)) = \phi_n(n^{1+\varepsilon}) \sim n^{1-\varepsilon} - (n^{1+\varepsilon} + n + 1) + (n + n^{1-\varepsilon} + 1/n^\varepsilon) \sim n^{1-\varepsilon}.$$

Hence, from (3.1) we have that $\mathbb{E}(X_{n-1:n}) \leq \mathbb{E}(X_1)\max\{n^{1-\varepsilon}, n\Delta_n\}$, and dividing by $\frac{n}{\log(\log(n))}$ we get

$$\frac{\mathbb{E}(X_{n-1:n})}{\frac{n}{\log(\log(n))}} \leq \mathbb{E}(X_1)\max\left\{\frac{\log(\log(n))}{n^\varepsilon}, \log(\log(n))\Delta_n\right\}$$

and taking liminf we conclude the statement. □

3.4 Lower bound

We adapt the argument made for the impossibility result of the maximum in the previous chapter. The following straightforward result summarizes the general methodology to obtain a lower bound.

Lemma 3.7 *Suppose there exist functions $N_1, N_2, f : \mathbb{N} \to \mathbb{R}_+$ and $\varepsilon > 0$ such that for all sufficiently large n, $\phi_n(x) \geq \varepsilon f(n)$ for all $x \in [N_1(n), N_2(n)]$. Suppose also that there exists a positive sequence $(\delta_k)_k \subseteq \mathbb{R}_+$ such that $\sum \delta_k/k < \infty$ and*

$$\sum_{k=N_1(n)}^{N_2(n)} \frac{\delta_k}{k} = \Omega(g(n)),$$

for some function $g : \mathbb{N} \to \mathbb{R}_+$. Then, $\mathbb{E}(X_{n-1:n}) = \Omega(f(n)g(n))$.

[1] Indeed, suppose by contradiction that there exists $\eta > 0$ and $N \in \mathbb{N}$ such that $na_n \geq \eta$ for all $n > N$. Then, $\sum_{n>N} a_n \geq \sum_{n>N} \frac{\varepsilon}{n} = \infty$.

Now we use Lemma 3.7 to derive a simple lower bound, and leave the improvement of this bound as an open question.

Theorem 3.8 *There exists a distribution F with finite expectation such that if $X_1, \ldots, X_n$ are independently drawn from F, then*

$$\mathbb{E}(X_{n-1:n}) = \Omega\left(\frac{n}{\log(n)^2}\right).$$

Proof. From Proposition 3.4 we know that $\phi_n(x) \geq \phi_n(2n)$, for $x \in [n, 2n]$. Furthermore, for sufficiently large n

$$\phi_n(2n) = 2n - \left(1 - \frac{1/2}{n}\right)^{n-1}(3n-1) \geq 2n - \mathrm{e}^{-1/2}(3n-1) = (2 - 3\mathrm{e}^{-1/2})n - 1 \geq \varepsilon n,$$

for some $0 < \varepsilon < 2 - 3\mathrm{e}^{-1/2} \approx 0.18$. Now, choosing

$$\delta_k = \left(\frac{1}{\log(k)} - \frac{1}{\log(k+1)}\right)k$$

we get that $\sum \delta_k/k = \lim_{k\to\infty} 1/\log(k) = 0$, and

$$\sum_{k=n}^{2n} \frac{\delta_k}{k} = \frac{1}{\log(n)} - \frac{1}{\log(2n)} = \frac{\log(2)}{\log(n)(\log(n) + \log(2))} \geq \eta \frac{1}{(\log(n))^2},$$

for some $\eta > 0$ and all sufficiently large n. Thus, the statement follows from Lemma 3.7. $\square$

Chapter 4

Bibliography

[1] Barry C Arnold. p-norm bounds on the expectation of the maximum of a possibly dependent sample. *Journal of multivariate analysis*, 17(3):316–332, 1985.

[2] Barry C Arnold, Narayanaswamy Balakrishnan, and Haikady Navada Nagaraja. *A first course in order statistics*. SIAM, 2008.

[3] Vladimir Igorevich Bogachev and Maria Aparecida Soares Ruas. *Measure theory*, volume 1. Springer, 2007.

[4] Jose Correa, Patricio Foncea, Ruben Hoeksma, Tim Oosterwijk, and Tjark Vredeveld. Recent developments in prophet inequalities. *ACM SIGecom Exchanges*, 17(1):61–70, 2019.

[5] Herbert A David and Haikady N Nagaraja. *Order statistics*. John Wiley & Sons, 2004.

[6] Peter J Downey. Distribution-free bounds on the expectation of the maximum with scheduling applications. *Operations Research Letters*, 9(3):189–201, 1990.

[7] Soheil Ehsani, MohammadTaghi Hajiaghayi, Thomas Kesselheim, and Sahil Singla. Prophet secretary for combinatorial auctions and matroids. In *Proceedings of the twenty-ninth annual acm-siam symposium on discrete algorithms*, pages 700–714. SIAM, 2018.

[8] David Freedman. *Markov chains*. Springer Science & Business Media, 2012.

[9] Janos Galambos. The asymptotic theory of extreme order statistics. Technical report, 1978.

[10] Xiaohu Li and Rui Fang. Ordering properties of order statistics from random variables of archimedean copulas with applications. *Journal of Multivariate Analysis*, 133:304–320, 2015.

[11] Marian Muresan and Marian Muresan. *A concrete approach to classical analysis*, volume 14. Springer, 2009.