Scaling of the Minimum of iid Random Variables

T. Y. Al-Naffouri*

Electrical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Abstract

Several areas in signal processing and communications rely on various tools in order statistics. Studying the scaling of the extreme values of iid random variables is of particular interest as it is sometimes only possible to make meaningful statements in the large number of variables case. This paper develops a new approach to finding the scaling of the minimum of iid variables by studying the behavior of the CDF and its derivatives at one point, or equivalently by studying the behavior of the characteristic function. The theory developed is used to study the scaling of several types of random variables and is confirmed by simulations.

Keywords: scaling of random variables — extreme values — order statistics— characteristic function — initial value theorem

1 Introduction

Extreme value theory (EVT) is an important tool in statistics, signal processing, and communications. EVT is concerned with the behavior of the maximum/minimum of a sequence of n iid random variables when n becomes large (which is known as scaling analysis). This tool has for example been used in abnormality detection in biomedical signal processing [1], for level detection in hidden Markov models [2], and in characterization of sonar reverberations [3]. ETV

* *E-mail addresse:* `naffouri@kfupm.edu.sa`

has also found wide application in multiuser information theory where it was used to schedule users, maximize system throughput, and study fairness issues, ..., etc [5, 4, 6, 7].

Performing the scaling analysis in turn requires closed form expressions for the CDF and pdf of the variables involved [9, 8] which might still be too prohibitive. Thus, at times we don't have closed expressions for the pdf or CDF at all points of support or these expressions might only be available at specific points. Moreover, sometimes it might be easier to characterize the behavior of the characteristic function instead.

In this paper, we show how the scaling law of the minimum of iid random variables can be obtained by studying the behavior of the CDF and its derivatives at one point. We also show how this can be obtained by studying the behavior of the characteristic function at infinity.

2 What Does Scaling Mean?

Let $X_1, X_2, \cdots, X_n$ be iid random variables with pdf $f(x)$, CDF $F(x)$, and characteristic function $\phi(s)$. Let a also be the infimum of the support of X_i[1] We would like to find the scaling law of the minimum, $X_{\min}(n) = \{X_1, X_2, \cdots, X_n\}$. We say that a variable scales for large n if there are sequences a_n and b_n such that $F^n(a_n X + b_n) \to G(x)$ as $n \to \infty$ at all continuity points of $G(x)$. It has been shown that when such a $G(x)$ exists, it falls into one of three categories [9]:

$$\begin{aligned}
\text{(Fréchet)} \quad G_1(x;\alpha) &= \exp(-x^{-\alpha})u(x) \quad \alpha > 0 \\
\text{(Weibull)} \quad G_2(x;\alpha) &= \exp(-(-x)^{\alpha})u(-x) \quad \alpha < 0 \\
\text{(Gumbel)} \quad G_3(x) &= \exp(-e^{-x})
\end{aligned} \tag{1}$$

Similar asymptotic distributions exist for the minimum of iid random variables. Specifically, there are 3 limiting distributions [9] $G_i^*(x)$ defined in terms of the maximum pdf counterpart $G_i^*(x) = 1 - G_i(-x)$. We illustrate this definition with an example.

[1] If no finite infimum exists, we can obtain the scaling of the minimum by considering the scaling of $-\max\{-X_1, -X_2, \cdots, -X_n\}$.

Example: Beamforming for multicast In a multicast beamforming scenario, the transmitter has M antennas and is to multicast some *common* data to a group of n users. The transmission rate is eventually limited by the worst user. The transmitter sends M beams $\phi_1, \phi_2, \ldots, \phi_M$, and asks each user to feedback the SINR associated with each beam [4]. For example the SINR associated with beam ϕ_1 for user i is given by

$$\mathrm{SINR}_i = \frac{|h_i^* \phi_1|^2}{\frac{1}{\rho} + \sum_{m=2}^{M} |h_i^* \phi_m|^2} \tag{2}$$

where h_i is the channel impulse response for user i and $\rho = \frac{M}{P}$ is the signal to noise ratio. Since the base station has to appeal to all users in the group, we are constrained by the worst user. For a given beam, the SINR_i's are iid with CDF [4]

$$F_{\mathrm{SINR}}(x) = 1 - \frac{e^{-\frac{x}{\rho}}}{(1+x)^{M-1}} \quad x \geq 0$$

So the CDF of the minimum of n such SINR's is

$$F_{\min}(x) \;=\; 1 - (1 - F_{\mathrm{SINR}}(x))^n = 1 - \frac{e^{-\frac{nx}{\rho}}}{(1+x)^{n(M-1)}} \quad x \geq 0 \tag{3}$$

Now, it is easy to show that

$$\lim_{n \to \infty} F_{\min}\left(\frac{x}{(\frac{1}{\rho} + M - 1)n}\right) \;=\; 1 - e^{-x} = 1 - G_2(-x; 1)$$

This shows that for large n, $\min_i \mathrm{SINR}_i$ scales as $\frac{1}{(\frac{1}{\rho}+M-1)n}$. The method of Example 1 might not apply all the time as it is difficult to find the CDF in closed form sometimes.

Example 2: Scaling of spatially correlated channel norms Consider the issue of finding the scaling of the $\min_i \|h_i\|^2$ where h_i is circularly symmetric Gaussian distributed $h_i \sim \mathcal{N}(m, R)$. When $R = I$ and $m = 0$, $\|h_i\|^2$ is chi-square distributed with M degrees of freedom. On the other hand, when we deviate from this ideal case, say when $R \neq I$, the CDF has different forms depending on whether some of the eigenvalues of the R are the same or not. In the case that $m \neq 0$, we don't even have a closed expression for the CDF (or pdf). As such, it would be difficult to find the scaling of the minimum using available techniques. In the following subsection, we provide a general method for finding the scaling of the minimum.

3 Evaluating the Scaling Using the Characteristic Function

In the following section, we demonstrate how the scaling of the minimum can be obtained directly from the characteristic function. To this end, note that the CDF of the minimum is given by $F_{\min}(x) = 1 - (1 - F(x))^n$. Let's expand $F(x)$ in a Taylor series

$$F(x) \quad = \quad \sum_{i=0}^{\infty} F^{(i)}(a) \frac{(x-a)^i}{i!} \tag{4}$$

Note that $F(a) = 0$ and let $F_{\min}^{(i_0)}(a)$ be the first nonzero derivative of $F(x)$ around a. Then

$$F_{\min}(x) = 1 - (1 - \frac{F^{(i_0)}(a)}{i_0!}(x-a)^{i_0} - \sum_{i \geq i_0} F^{(i)}(a) \frac{(x-a)^i}{i!})^n \tag{5}$$

Now, we claim that $F^{(i_0)}(a) > 0$. For if it were negative, then $F^{(i_0-1)}(a)$ would be decreasing in an interval $(a, a+\epsilon)$. Or as $F^{(i_0)}(a) = 0$, we see that $F^{(i_0-1)}(x)$ is negative in this interval. Continuing this way, we can show that $F^{(i_0-1)}(x), \ldots, F^{(0)}(x)$ are negative in $(a, a+\epsilon)$ which contradicts the nonnegative nature of $F(x)$. We can thus replace x by $\frac{i_0!^{\frac{1}{i_0}}}{F^{(i_0)}(a)^{\frac{1}{i_0}}} \frac{x+a}{n^{\frac{1}{i_0}}}$. Then

$$F_{\min}(\frac{i_0!^{\frac{1}{i_0}}}{F^{(i_0)}(a)^{\frac{1}{i_0}}} \frac{x+a}{n^{\frac{1}{i_0}}}) = 1 - \left(1 - \frac{x^{i_0}}{n} + O(\frac{1}{n^{\frac{i_0+1}{i_0}}})\right)^n$$

which for large n reads

$$\lim_{n \to \infty} F_{\min}(\frac{i_0!^{\frac{1}{i_0}}}{F^{(i_0)}(a)^{\frac{1}{i_0}}} \frac{x+a}{n^{\frac{1}{i_0}}}) = 1 - \exp\left(-x^{i_0}\right) \tag{6}$$

This is of the form

$$\begin{aligned} \lim_{n \to \infty} F_{\min}(a_n x + b_n) \quad &= \quad 1 - \exp\left(-x^{i_0}\right) && (7) \\ &= \quad 1 - G_2(-x; i_0) && (8) \end{aligned}$$

where

$$a_n = \frac{i_0!^{\frac{1}{i_0}}}{F^{(i_0)}(a)^{\frac{1}{i_0}}} \frac{1}{n^{\frac{1}{i_0}}} \quad \text{and} \quad b_n = \frac{i_0!^{\frac{1}{i_0}}}{F^{(i_0)}(a)^{\frac{1}{i_0}}} \frac{a}{n^{\frac{1}{i_0}}} \tag{9}$$

3.1 Finding the derivatives of $F(a)$

The scaling above is determined above up to the first nonzero derivative of $F^{(i_0)}(a)$. Fortunately, we can find this value without having to explicitly find the CDF and its derivatives, by using the characteristic function and relying instead on the initial value theorem. To do so, define $D(x) = F(x+a)$, then $\lim_{x\to 0} D^{(j)}(x) = \lim_{x\to a} F^{(j)}(x)$. Now recall that the pdf $f(x)$ and the characteristic function $\phi(s)$ form a Laplace transform pair. Then by the time-shift and differentiation properties

$$D^{(j)}(x) \to s^{(j-1)} e^{-as}\phi(s)$$

Now applying the initial value theorem to the Laplace transform pair above yields

$$\lim_{x\to 0} D^{(j)}(x) = \lim_{s\to\infty} s^{j} e^{-as}\phi(s)$$

i.e.

$$\lim_{x\to a} F^{(j)}(x) = \lim_{s\to\infty} s^{(j-1)} e^{-as}\phi(s) \tag{10}$$

We can summarize the results of this section in the following theorem.

Proposition 1 *Let $X_1, X_2, \cdots, X_n$ be iid random variables with CDF $F(x)$, and characteristic function $\phi(s)$. Assume that X_i is bounded from below and let a be the infimum of the support of X_i. Let $X_{\min}(n)$ denote the minimum of these random variables $\min\{X_1, X_2, \cdots, X_n\}$. Then $a_n x_{\min}(n) + b_n$ converges in distribution to random variable y with CDF $F_y(y) = 1 - \exp\left(-y^{i_0}\right)$ where i_0 is the first non-zero derivative of $F(x)$ at zero, i.e., $F^{(i_0)}(a) \neq 0$ and $F^{(j)}(a) = 0$ for all $j < i_0$ and where a_n and b_n are defined in (9). Furthermore, we can find $F^{(i_0)}(a)$ using the initial value theorem, $\lim_{x\to a} F^{(j)}(x) = \lim_{s\to\infty} s^{(j-1)} e^{-as}\phi(s)$.*

3.2 When does the method fail?

The method introduced in this paper has wide applicability. However, as can be inferred from the paper, it fails if for all i, $F^{(i)}(a) = 0$. One example for which this is the case is the inverse chi-square pdf given by $f(x) = e^{-\frac{1}{x}}/x^2, x > 0$, for which $f(x)$ and all higher order derivatives are zero at $x = 0$.

4 Examples and Numerical Simulations

4.1 Example: Scaling of spatially correlated channel norms revisited

Let's find the scaling law for $\min_{h_i} \|h_i\|^2$ when h_i are iid $CN(0, R)$. The pdf and CDF of $\|h_i\|^2$ will both have different forms depending on whether some of the eigenvalues λ_l of R are the same or different, and so the direct method for scaling can be quite challenging. On the other hand, the characteristic function takes one form and is given by $\phi(s) = \prod_{l=1}^{M} \frac{1}{1+\lambda_l s}$. From this, it is easy to see that

$$\lim_{s\to\infty} s^i \phi(s) = F^{(i)}(0) = 0 \ \text{ for } i < M$$

and that $\lim_{s\to\infty} s^M \phi(s) = F^{(M)}(0) = \frac{1}{\prod_{l=1}^{M} \lambda_l} = \frac{1}{\det(R)}$. We thus conclude that $\min_i \|h_i\|^2$ scales as $(M!)^{\frac{1}{M}} \det(R)^{\frac{1}{M}} \frac{1}{n^{\frac{1}{M}}}$.

4.2 Examples for the scaling of minimum of a number of random variables

In the following, we evaluate the scaling of various random variables and summarize the results in the Table 1 below. The table gives the name of the distribution, expressions for pdf, CDF, and characteristic function whenever they are available. The table also provides the values of a, i_0, and $F^{(i_0)}(a)$ which is the information needed to characterize the scaling behavior. For some of these distributions (namely, Uniform, non-central chi-square, Gamma, and Half normal distributions), we simulate the empirical and theoretical CDF of the minimum as well as the actual value of the minimum random variable vs. n (averaged over 300 runs) and the theoretical scaling value predicted. The figures show very good match between theory and simulation.

4.2.1 Uniform distribution

Let x be a Uniform random variable, $x \in (c, d)$. In this simulation (Figure 1) we take $c = 3$ and $d = 6$. In this case, we have $a = c$, $i_0 = 1$, and $F^{(i_0)}(a) = \frac{1}{d-c}$.

4.2.2 Non-central chi-square distribution

Let x be a non-central chi-square random variable with 4 degress of freedom (resulting from the sum of squares of 4 Gaussian random variables with mean $m = 1$). In this case, the CDF has no closed form. However, using the characteristic function approach, we can show that $a = 0$, $i_0 = 1$, and $F^{i_0}(0) = 1/4\, e^{-2\,\frac{m_x{}^2}{\sigma^2}}\,\sigma^{-4}$. The CDF and the plot of X_{min} vs n are shown in Figure 2.

4.2.3 Gamma distribution

Figure 3 shows the distribution and scaling of the minimum of gamma random variables. This random variable has the characteristic function $(1-\theta s)^{-k}$ so that $a = 0$, $i_0 = k$ and $F^{i_0}(0) = \theta^{-k}$.

4.2.4 Half normal distribution

Figure 4 shows the distribution and scaling of the minimum of half-nromal random variables. In this case, $a = 0$, $i_0 = 1$ and $F^{i_0}(0) = \frac{1}{\sigma}\sqrt{\frac{2}{\pi}}$.

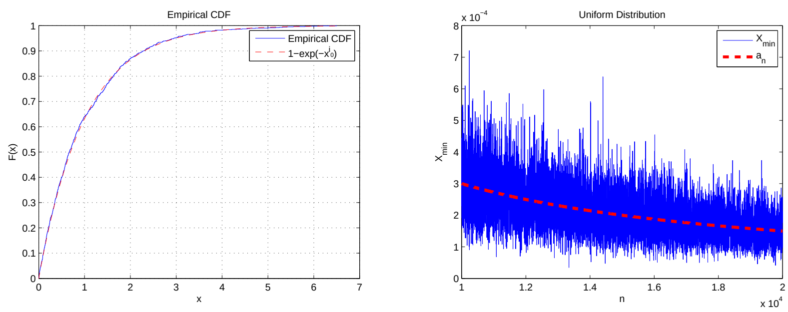

Figure 1: Uniform Distribution

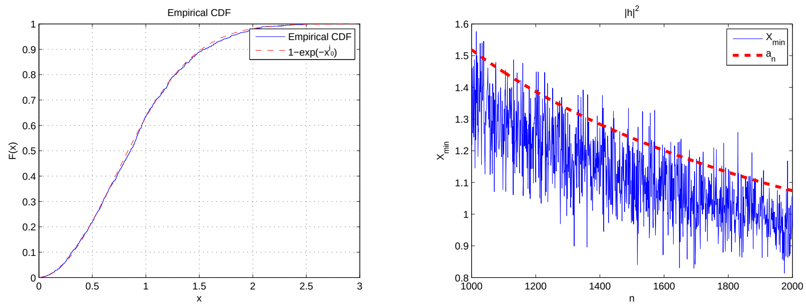

Figure 2: Non-Central Chi Square Distribution

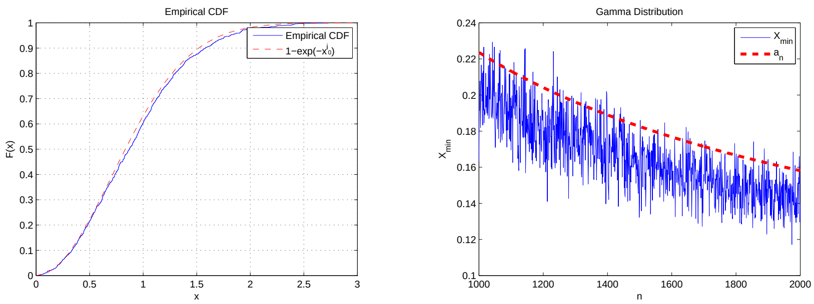

Figure 3: Gamma Distribution

5 Conclusion

In this paper, we devised a new method for characterizing the scaling of the minimum of iid random variables. The method is based on studying the behavior of the higher order derivatives of the associated CDF at the infimum of the random variable. Equivalently, the scaling can be studied directly from the characteristic function. The method thus circumvents the need to have explicit expressions for the CDF of pdf to study the scaling. The method was used to characterize the scaling of several random variables. Theoretical results showed very good match to our simulations. The author is currently expanding this approach to characterize the scaling of the maximum of iid random variables.

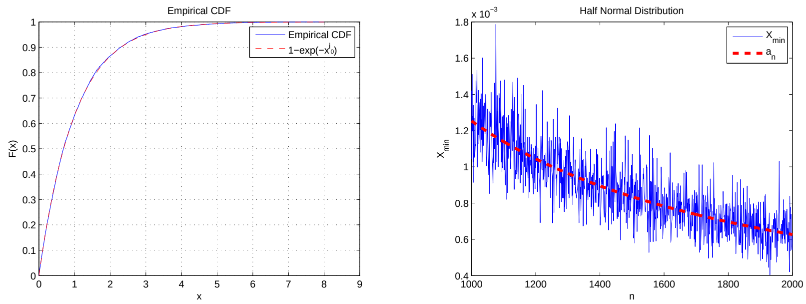

Figure 4: Half Normal Distribution

References

[1] S. Roberts, "Extreme value statistics for novelty detection in biomedical signal processing," *International Conference on Advances in Medical Signal and Information Processing,* Sep. 2000.

[2] K. Dogancay and V. Krishnamurthy, "Application of extreme value theory to level estimation in nonlinearly distorted hidden Markov models," *IEEE Trans. Sig. Proc.,* no. 8, vol 48, Aug. 2000.

[3] B.R. La Cour, "Statisticl chracterization of active sonar reverberation using extreme value theory," *IEEE Journal of Oceanic Engineering,* no. 2, vol. 29, Apr. 2004.

[4] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Info. Theory*, vol. 51, no. 2, Feb. 2005.

[5] P. Viswanath, D. N. Tse, and R. Laroia, "Opportunistic beamforming using dump antennas," *IEEE Trans. Inform.*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[6] T. Y. Al-Naffouri, M. Sharif, and B. Hassibi, "How Much Does Transmit Correlation Affect the Sum-Rate of MIMO Downlink Channels?" *to appear in IEEE Transactions. on Communications.*

[7] T. Y. Al-Naffouri, A. Dana, and B. Hassibi, "Scaling laws of group broadcast channels," *to appear in IEEE Transactions on Wireless Communications.*

[8] N. T. Uzgoren, "The asymptotic development of the distribution of the extreme values of a sample," *Studies in Mathematics and Mechanics Presented to Richard von Mises,* Academic Press, New York, pp. 346–353, 1954.

[9] H. A. David, *Order Statistics*, New York, Wiley, 1970.

Table 1:

	$p(x)$	$P(x)$	$\phi(s)$	a	i_0	$F^{i_0}(0)$
Uniform Dist.	$\frac{1}{d-c}$			c	1	$\frac{1}{d-c}$
U-Quatratic Dist.	$\alpha(x-\beta)^2$	$\frac{\alpha}{3}((x-\beta)^3+(\beta-\alpha)^3)$		$\beta - \sqrt[3]{\frac{3}{2\alpha}}$	1	$\sqrt[3]{\frac{9\alpha}{4}}$
Non-Central Chi Square Dist.			$\frac{1}{(1-2\sigma^2 s)^{n/2}} e^{\frac{\sum_{i=1}^n m_i^2}{1-2\sigma^2 s}}$	0	$\frac{n}{2}$	$\frac{1}{2^{n/2}\sigma^n} e^{\frac{-\sum_{i=1}^n m_i^2}{2\sigma^2}}$
Chi- Square Dist.			$\frac{1}{(1-2\sigma^2 s)^{n/2}}$	0	$\frac{n}{2}$	$\frac{1}{2^{n/2}\sigma^2}$
Gamma Dist.	$x^{k-1}\frac{e^{-x/\theta}}{\Gamma(k)\theta^k}$	$\frac{\gamma(k,x/\theta)}{\Gamma(k)}$	$(1-\theta s)^{-k}$	0	k	θ^{-k}
Rayleigh Dist.	$\frac{xe^{-x^2/2\sigma^2}}{\sigma^2}$	$1-e^{-x^2/2\sigma^2}$		0	2	$\frac{1}{\sigma^2}$
Pareto Dist.	$\frac{kx_m^k}{x^{k+1}}$	$1-(\frac{x_m}{x})^k$		1	1	$\frac{k}{x_m}$
Log-Logistic Dist.	$\frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{(1+(x/\alpha)^\beta)^2}$	$\frac{1}{1+(x/\alpha)^{-\beta}}$		0	1	1
Half Normal Dist.		$\int_0^x \frac{1}{\sigma}\sqrt{\frac{2}{\pi}} e^{-y^2/2\sigma^2} dy$		0	1	$\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}$
Folded Normal Dist.		$\int_0^x \frac{1}{\sigma\sqrt{2\pi}} e^{-(-y-\mu)^2/2\sigma^2} dy +$ $\int_0^x \frac{1}{\sigma\sqrt{2\pi}} e^{(-y-\mu)^2/2\sigma^2} dy$		0	1	$\frac{1}{\sigma}\sqrt{\frac{2}{\pi}} e^{-\mu^2/2\sigma^2}$
Kumaraswamy Dist.	$cdx^{c-1}(1-x^c)^{d-1}$			0	c	$cd(c-1)!$
Beta Dist.	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$			0	α	$\frac{(\alpha-1)!}{B(\alpha,\beta)}$