Inequalities relating maximal moments to other measures of dispersion

Pieter C. Allaart

March 31, 2000

University of North Texas

Abstract

Let $X, X_1, \ldots, X_k$ be i.i.d. random variables, and for $k \in \mathbb{N}$ let $D_k(X) = E(X_1 \vee \cdots \vee X_{k+1}) - EX$ be the k-th centralized maximal moment. A sharp lower bound is given for $D_1(X)$ in terms of the Lévy concentration $Q_l(X) = \sup_{x \in \mathbb{R}} P(X \in [x, x+l])$. This inequality, which is analogous to P. Lévy's concentration-variance inequality, illustrates the fact that maximal moments are a gauge of how much spread out the underlying distribution is. It is also shown that the centralized maximal moments are increased under convolution.

AMS 1990 subject classifications: 60E15, 60G70.
Keywords and phrases: expected maximum of an i.i.d. random sample, Lévy concentration, measure of dispersion.

Author's address: Mathematics Department, University of North Texas, Denton, TX 76203-1430, USA; e-mail: allaart@unt.edu

1 Introduction

Let $X, X_1, \ldots, X_k$ be independent, identically distributed random variables with finite mean and common distribution function F, and let

$$\begin{aligned} M_k := M_k(X) &:= E\max\{X_1, \ldots, X_k\} \\ &= \int x dF^k(x) \end{aligned} \tag{1}$$

denote the expected value of the largest order statistic. Such expectations, to be called *maximal moments*, have been studied for several decades, a sizeable part of the literature consisting of bounds for maximal moments in terms of other characteristics of the underlying distribution, such as the variance or, more generally, the p-th absolute moment (cf. Arnold, 1985, and Hartley and David, 1981). Few authors, however, have emphasized that the maximal moments themselves, or rather the differences

$$D_k := D_k(X) := M_{k+1}(X) - M_1(X) = M_{k+1}(X) - EX$$

can serve as a gauge of how spread out the distribution is. For instance, $D_k = 0$ for some $k \geq 1$ if and only if X is degenerate; the D_k's are translation-invariant, that is, $D_k(X + b) = D_k(X)$; and they are linear: $D_k(aX) = aD_k(X)$ for every $a > 0$. Moreover, D_1 has the additional desirable property of being symmetric, that is, $D_1(-X) = D_1(X)$. An obvious advantage of maximal moments over for instance variance is, that they do not require a finite second moment: all are automatically finite when $E|X|$ is finite. A possible drawback is that their exact values are sometimes difficult, if not impossible, to compute. For the standard normal distribution $N(0, 1)$, for example, only the first five are known exactly, though numeric approximations of higher order maximal moments are available; cf. David, 1981.

The aim of the present note is two-fold: first it will be shown that D_k is increased by convolution, in correspondence with the intuitive idea that the distribution of a sum must be more spread out than each of the component distributions. This is done in Section 2.

Section 3 contains the main result of this paper, a sharp lower bound for D_1 in terms of Lévy concentration analogous to P. Lévy's concentration-variance inequality. For completeness, bounds in terms of the variance and the first absolute moment about the mean are also given. Section 4 contains a direct application of the main result to the linear search problem.

2 D_k is increased by convolution

Throughout this paper, the following notation will be used: $x \vee y$ denotes the maximum of x and y, $x^+ = x \vee 0$ and $x^- = (-x) \vee 0$ denote the positive, resp. negative part of x, $\lfloor x \rfloor$ is the greatest integer less than or equal to x, and for distributions P and Q, $P * Q$ denotes the convolution of P and Q.

Among other things already summed up in the introduction, one would expect a candidate measure of dispersion to behave predictably under convolution of the underlying distribution with another distribution. To illustrate, just consider the variance identity $\sigma^2_{P_1 * P_2} = \sigma^2_{P_1} + \sigma^2_{P_2}$. Another example is the *Lévy concentration*

$$Q_l(P) = \sup_{x \in \mathbb{R}} P([x, x+l]); \tag{2}$$

it was shown by P. LÉVY (1937) that for any two distributions P_1 and P_2

$$Q_l(P_1 * P_2) \leq \min\{Q_l(P_1), Q_l(P_2)\}. \tag{3}$$

The following proposition says that the centralized maximal moments D_k satisfy an analogous inequality:

Proposition 2.1 *Let P and Q be probability distributions, and $k \geq 1$. Then*

$$D_k(P) \vee D_k(Q) \leq D_k(P * Q) \leq D_k(P) + D_k(Q). \tag{4}$$

Both inequalities are attained if and only if P or Q is degenerate.

Proof. Note first that if, for instance, Q is degenerate, then $D_k(Q) = 0$ and hence both parts of (4) hold with equality. Suppose therefore that both P and Q are non-degenerate. To prove strict inequality, we will use the basic principle that

$$\text{if} \quad W \geq Z \quad a.s. \quad \text{and} \quad Pr(W > Z) > 0, \quad \text{then} \quad EW > EZ. \tag{5}$$

Let $X_1, \ldots, X_{k+1}$ be i.i.d. random variables with distribution P, and let $Y_1, \ldots, Y_{k+1}$ be i.i.d. with distribution Q, such that $(X_1, \ldots, X_{k+1})$ is independent of $(Y_1, \ldots, Y_{k+1})$. Observe that

$$D_k(P) = E(X_1 \vee \cdots \vee X_{k+1}) - EX_1 = E[(X_2 - X_1) \vee \cdots \vee (X_{k+1} - X_1)]^+,$$

and similarly,

$$\begin{aligned} D_k(P * Q) &= E[(X_1 + Y_1) \vee \cdots \vee (X_{k+1} + Y_{k+1})] - E(X_1 + Y_1) \\ &= E[(X_2 - X_1 + Y_2 - Y_1) \vee \cdots \vee (X_{k+1} - X_1 + Y_{k+1} - Y_1)]^+. \end{aligned}$$

Define $U_i := X_{i+1} - X_1$ and $V_i := Y_{i+1} - Y_1$. Then $D_k(P * Q) < D_k(P) + D_k(Q)$ is equivalent to

$$E[(U_1 + V_1) \vee \cdots \vee (U_k + V_k)]^+ < E(U_1 \vee \cdots \vee U_k)^+ + E(V_1 \vee \cdots \vee V_k)^+. \tag{6}$$

To prove (6), note that for all real numbers $u_1, \dots, u_k, v_1, \dots, v_k$, $[(u_1+v_1)\vee\cdots\vee(u_k+v_k)]^+ \le (u_1 \vee \cdots \vee u_k)^+ + (v_1 \vee \cdots \vee v_k)^+$. Since strict inequality holds for example when $u_i < 0$ and $v_i > 0$ for all i, it is sufficient to show that $Pr(U_1 < 0, \dots, U_k < 0, V_1 > 0, \dots, V_k > 0) > 0$. By independence of $(X_1, \dots, X_{k+1})$ and $(Y_1, \dots, Y_{k+1})$, this probability is equal to $Pr(U_1 < 0, \dots, U_k < 0) \cdot Pr(V_1 > 0, \dots, V_k > 0)$. Both probabilities are positive for similar reasons, e.g. $Pr(U_1 < 0, \dots, U_k < 0) = Pr(X_2 < X_1, \dots, X_{k+1} < X_1) \ge Pr(X_1 > 0, X_2 \le 0, \dots, X_{k+1} \le 0) = Pr(X_1 > 0)Pr(X_2 \le 0) \cdots Pr(X_{k+1} \le 0) > 0$ by the non-degeneracy of P. An appeal to (5) yields (6).

To prove the other strict inequality, note that $D_k(P * Q) > D_k(P)$ is equivalent to

$$E[(U_1 + V_1) \vee \cdots \vee (U_k + V_k)]^+ \ge E(V_1 \vee \cdots \vee V_k)^+. \tag{7}$$

Observe that since U_i is non-degenerate and $EU_i = 0$, there exists $\varepsilon > 0$ such that $Pr(U_i < -\varepsilon) > 0$. To prove (7), we first show that for all $v_1, \dots, v_k$,

$$E[(U_1 + v_1) \vee \cdots \vee (U_k + v_k)]^+ \ge (v_1 \vee \cdots \vee v_k)^+, \tag{8}$$

with strict inequality if $0 \le v_1 \vee \cdots \vee v_k < \varepsilon$. Fix i such that $v_i = v_1 \vee \cdots \vee v_k$. Then

$$\begin{aligned} E[(U_1 + v_1) \vee \ldots \vee (U_k + v_k)]^+ &\ge E(U_i + v_i)^+ \\ &\ge [E(U_i + v_i)]^+ = v_i^+ = (v_1 \vee \cdots \vee v_k)^+. \end{aligned}$$

Here the second inequality is strict if $0 \le v_1 \vee \cdots \vee v_k < \varepsilon$, because $Pr(U_i + v_i < 0) \ge Pr(U_i < -\varepsilon) > 0$. This proves (8), and the corresponding strict inequality. To complete the proof of the Proposition, choose a real number a such that $Pr(a < Y_i \le a+\varepsilon) > 0$. Then $Pr(0 \le V_1 \vee \cdots \vee V_k < \varepsilon) \ge Pr(a < Y_1 \le Y_2 \vee \cdots \vee Y_{k+1} \le a+\varepsilon) > 0$ by i.i.d.-ness of $Y_1, \dots, Y_{k+1}$. Apply (5). $\square$

The first inequality in (4) may fail if D_k is replaced by the *consecutive* maximal moment difference $d_k := M_{k+1} - M_k$. Notice that $d_1 = D_1$, $d_1 + d_2 = D_2$, etc. The next example illustrates what can go wrong if $k \ge 2$.

Example 2.2 Let $P = \frac{3}{4}\delta_{\{0\}} + \frac{1}{4}\delta_{\{1\}}$ and $Q = \frac{1}{10}\delta_{\{0\}} + \frac{9}{10}\delta_{\{1\}}$, where $\delta_{\{x\}}$ denotes Dirac measure at x. Then $P * Q = \frac{3}{40}\delta_{\{0\}} + \frac{7}{10}\delta_{\{1\}} + \frac{9}{40}\delta_{\{2\}}$, and a straightforward calculation shows that $d_2(P * Q) = 9 \cdot 998/64000 < 9/64 = d_2(P)$.

3 Bounds in terms of other measures of spread

In this section sharp upper and lower bounds are given for centralized maximal moments in terms of other measures of dispersion, including the standard deviation

σ, the first absolute moment about the mean $E|X-EX|$, and the Lévy concentration (2). The following sharp inequality is due to HARTLEY and DAVID (1981):

$$D_k \leq k(2k+1)^{-1/2}\sigma. \tag{9}$$

In particular,

$$D_1 \leq \sigma/\sqrt{3}. \tag{10}$$

Equality holds in (9) if and only if X is degenerate (then $D_1 = \sigma = 0$) or X has a distribution function G in the location-scale family of $F(x) = x^{1/k}$ for $0 \leq x \leq 1$. (For a proof of this fact, see the proof of Theorem 3 in ARNOLD (1985), and note that D_1/σ is invariant under change of scale or location.) In particular, equality holds in (10) if and only if X is degenerate or X is uniformly distributed.

Generalizations of the above inequalities were obtained by ARNOLD (1985), who has sharp upper bounds for D_k in terms of the p-th central moment $E|X-EX|^p$ for $p > 1$. In case $p = 1$, we have the following:

$$D_1 \leq E|X - EX| \leq 2D_1. \tag{11}$$

To see (11), note that by symmetry $2D_1 = 2E(X_2 - X_1)^+ = E|X_2 - X_1|$. The first inequality in (11) now follows since $2D_1 = E|X_2 - EX_2 - (X_1 - EX_1)| \leq 2E|X - EX|$. The second inequality can be proved as follows:

$$\begin{aligned} D_1 = E(X_2 - X_1)^+ &= \int E[(X_2 - X_1)^+ | X_1 = x] \; dP_{X_1}(x) \\ &= \int E(X_2 - x)^+ \; dP_{X_1}(x) \geq \int (EX_2 - x)^+ \; dP_{X_1}(x) \\ &= E(EX_2 - x)^+ = (1/2)E|X - EX|. \end{aligned}$$

Here the inequality follows by applying Jensen's inequality for the convex function $\phi_x(t) = (t-x)^+$, and the last equality follows using the algebraic identities $a^+ - a^- = a$ and $a^+ + a^- = |a|$, and the fact that $X - EX$ has mean zero.

The constant 2 in (11) is best possible, and equality holds for any two-point distribution.

The following theorem is the main result of this paper.

Theorem 3.1 *For each probability distribution P and for each $l > 0$,*

$$D_1(P) \geq \frac{l}{6} n(n+1) Q_l(P) \{3 - (2n+1) Q_l(P)\}, \tag{12}$$

where $n = \lfloor 1/Q_l(P) \rfloor$, and this bound is sharp.

The above lower bound may remind the reader of P. Lévy's concentration-variance inequality

$$\sigma_P^2 \geq \frac{l^2}{12} n(n+1)\{3-(2n+1)Q_l(P)\}, \tag{13}$$

where $n = \lfloor 1/Q_l(P) \rfloor$. In fact, the extremal distributions for (12) and (13) are the same (see the measures $P_{\lambda,l}$ defined in the proof below), and the first part of the proof of Theorem 3.1 will go along the same lines as the proof of (13) (cf. FOLEY *et al.*, 1990).

Definition 3.2 *For a distribution P and a Borel set $A \subset \mathbb{R}$ with $P(A) > 0$, the* P-barycenter *of A is $(P(A))^{-1} \int_A x dP(x)$.*

Proof of Theorem 3.1. Fix P, fix $\varepsilon > 0$ and $l > 0$ and let $\lambda := Q_l(P)$. Let $n = \lfloor 1/\lambda \rfloor$, and define numbers $(\beta_k)_{k=0}^{2n}$ by

$$\beta_k = \begin{cases} 1 - n\lambda, & k \text{ even} \\ (n+1)\lambda - 1, & k \text{ odd.} \end{cases}$$

Let $s_{-1} = 0$ and $s_k = \beta_0 + \beta_1 + \cdots + \beta_k$ for $k = 0, \ldots, 2n$. Note that $\beta_{2n-k} = \beta_k$, and hence $s_{2n-k} = 1 - s_{k-1}$ $(k = 0, \ldots, 2n)$. Moreover,

$$s_{2j} = 1-(n-j)\lambda, \qquad s_{2j-1} = j\lambda \qquad (j = 0, \ldots, n). \tag{14}$$

For $k = 0, \ldots, 2n$, define $\hat{z}_k = kl/2$, and let $P_{\lambda,l}$ be the measure

$$P_{\lambda,l} = \sum_{k=0}^{2n} \beta_k \delta_{\{\hat{z}_k\}}.$$

Let $\mathcal{M}_\lambda$ be the set of all probability distributions $\hat{P}$ with $Q_l(\hat{P}) \leq \lambda$. There exists $P_1 \in \mathcal{M}_\lambda$ such that $D_1(P_1) \leq D_1(P) + \varepsilon$, P_1 is absolutely continuous with positive density everywhere, and $\int |x| dP_1(x) < \infty$. (To see this, let Q be a distribution with everywhere continuous density and $D_1(Q) \leq \varepsilon$ (one such Q is the normal distribution with mean 0 and variance ε^2, in view of (10)). Define $P_1 := P * Q$. Then Proposition 2.1 implies $D_1(P_1) \leq D_1(P) + \varepsilon$, and (3) implies $P_1 \in \mathcal{M}_\lambda$.)

Next, define $-\infty = r_0 \leq r_1 \leq \cdots \leq r_{2n+1} = \infty$ by $P_1((r_0, r_1)) = \beta_0, P_1([r_1, r_2)) = \beta_1, \ldots, P_1([r_{2n}, r_{2n+1})) = \beta_{2n+1}$. Let $z_0, \ldots, z_{2n}$ be the P_1-barycenters of the intervals $(r_0, r_1), [r_1, r_2), \ldots, [r_{2n}, r_{2n+1})$ respectively, and define

$$P_2 = \sum_{k=0}^{2n} \beta_k \delta_{\{z_k\}}.$$

As in HENGARTNER and THEODORESCU (1973), p.28 it can be shown that $P_2 \in \mathcal{M}_\lambda$. Since D_1 is the integral of the convex function $\phi(x, y) = (x - y)^+$, the definition of P_2 and the conditional version of Jensen's inequality imply $D_1(P_1) \geq D_1(P_2)$. (An alternative to using Jensen's inequality is to observe that P_2 is a *fusion* of P_1 (see ELTON and HILL, 1992), and hence P_1 convexly dominates P_2.) It will now be shown that $D_1(P_2) \geq D_1(P_{\lambda,l})$.

Since $P_2 \in \mathcal{M}_\lambda$, we have $z_{k+1} - z_{k-1} > l$ for all $k = 1, \ldots, 2n - 1$, which by iteration yields

$$z_{2n-k} - z_k > (n - k)l, \quad k = 0, \ldots, n - 1. \tag{15}$$

Now

$$\begin{aligned} D_1(P_2) &= \sum_{j=0}^{2n} z_j(s_j^2 - s_{j-1}^2) - \sum_{j=0}^{2n} z_j\beta_j = \sum_{j=0}^{2n} z_j\beta_j(s_{j-1} + s_j - 1) \\ &= \sum_{j=0}^{n-1} [z_j\beta_j(s_{j-1} + s_j - 1) + z_{2n-j}\beta_{2n-j}(s_{2n-j-1} + s_{2n-j} - 1)] \\ &\qquad\qquad + z_n\beta_n(s_{n-1} + s_n - 1) \\ &= \sum_{j=0}^{n-1} [z_j\beta_j(s_{j-1} + s_j - 1) + z_{2n-j}\beta_j(1 - s_j - s_{j-1})] + 0 \\ &= \sum_{j=0}^{n-1} (z_{2n-j} - z_j)\, \beta_j\, (1 - s_j - s_{j-1}) \\ &\geq \sum_{j=0}^{n-1} (n - j)l\beta_j(1 - s_j - s_{j-1}) \\ &= \sum_{j=0}^{n-1} (\hat{z}_{2n-j} - \hat{z}_j)\, \beta_j\, (1 - s_j - s_{j-1}) = D_1(P_{\lambda,l}), \end{aligned}$$

where the fourth equality follows by symmetry of $(\beta_k)_{k=0}^{2n}$, and the inequality follows by (15) since $1 - s_j - s_{j-1} \geq 0$ for $j \leq n - 1$.

It remains to calculate $D_1(P_{\lambda,l})$. Let X_1 and X_2 be independent random variables with distribution $P_{\lambda,l}$. Using Fubini's theorem in the usual way we can write

$$EX_1 = \frac{l}{2}\sum_{k=0}^{2n}(1 - s_k), \qquad \text{and} \qquad E(X_1 \vee X_2) = \frac{l}{2}\sum_{k=0}^{2n}(1 - s_k^2).$$

Thus

$$\begin{aligned} D_1(P_{\lambda,l}) = E(X_1 \vee X_2) - EX_1 &= \frac{l}{2}\sum_{k=0}^{2n} s_k(1-s_k) \\ &= \frac{l}{2}\sum_{j=0}^{n-1}(1-(n-j)\lambda)(n-j)\lambda + \frac{l}{2}\sum_{j=1}^{n} j\lambda(1-j\lambda) \\ &= \frac{l}{2}\sum_{i=1}^{n}(1-i\lambda)i\lambda + \frac{l}{2}\sum_{j=1}^{n} j\lambda(1-j\lambda) \\ &= l\sum_{j=1}^{n}\lambda j(1-\lambda j) = \frac{l}{6}n(n+1)\lambda(3-(2n+1)\lambda), \end{aligned}$$

where the third equality follows by (14). We now have

$$D_1(P) + \varepsilon \geq D_1(P_1) \geq D_1(P_2) \geq D_1(P_{\lambda,l}).$$

Letting $\varepsilon \downarrow 0$ completes the proof of (12). Although $P_{\lambda,l} \notin \mathcal{M}_\lambda$, taking distributions $\sum_{k=0}^{2n} \beta_k \delta_{\{y_k\}}$ with y_k arbitrarily close to the extremal case $kl/2$, and with $y_{k+1} - y_k > l/2$, shows the bound is sharp. □

Remark 3.3 If F is the distribution function corresponding to P, then using (1) and Fubini's theorem we can write

$$D_1(P) = \int_{-\infty}^{\infty} F(x)(1-F(x))\,dx.$$

Thus an alternative formulation of Theorem 3.1 is

$$\int_{-\infty}^{\infty} F(x)(1-F(x))\,dx \geq l\sum_{j=1}^{n}\lambda j(1-\lambda j), \tag{16}$$

where $\lambda = Q_l(P) = \sup_{x\in\mathbb{R}}(F(x+l) - F(x^-))$, and $n = \lfloor\lambda\rfloor$. Despite the striking similarity between the two expressions in (16) it is not clear how to give a direct analytical proof of (16), and hence of Theorem 3.1.

4 An application to the linear search problem

Beck (1964) describes the linear search problem as follows: "A man in an automobile searches for another man who is located at some point of a certain road. He starts at a given point and knows in advance the probability that the second man is at any given point of the road. Since the man being sought might be in either direction from the starting point, the searcher will, in general, have to turn around many times before finding his target. How does he search so as to minimize the expected distance traveled?"

In this section, we will apply Theorem 3.1 to a different, easier version of the problem. Suppose the searcher has only a limited amount of time, but is allowed to start the search wherever he wants. His purpose is to maximize the probability of finding the target within the allotted time. Clearly, if there is time for a search of length l, then this maximum probability is $Q_l(P)$, where P is the probability distribution on $\mathbb{R}$ that models the likelihood for the target to be at any given point of the road. The maximum is attained since there exists a point $x_0 \in \mathbb{R}$ such that $Q_l(P) = P([x_0, x_0 + l])$ (see HENGARTNER and THEODORESCU (1973), Theorem 1.1.8). Thus the optimal search plan is to start at x_0 and search l units to the right, and the probability of finding the target using this plan is $Q_l(P)$.

The following equivalent of Theorem 3.1 gives an explicit sharp lower bound for $Q_l(P)$ in terms of $D_1(P)$.

Theorem 3.1'

$$Q_l(P) \geq \frac{3}{2}(2n+1)^{-1}\left[1 + \left(1 - \frac{8(2n+1)D_1(P)}{3n(n+1)l}\right)^{-1/2}\right],$$

where n is the unique integer such that $(n^2-1)l \leq 6nD_1(P) < n(n+2)l$, and this bound is sharp.

Proof. Solve the inequality (12) for $Q_l(P)$, and observe that when $Q_l(P) = 1/n$, the lower bound in (12) is $l(n^2-1)/(6n)$.

Example 4.1 If an object is placed on the real line according to a distribution P with $D_1(P) = 2$, then there is a starting point x_0 and a search of length 3 starting at x_0 that will find the object with probability $\frac{1}{6}(1 + 1/\sqrt{5}) \approx 0.24$, or more.

References

[1] ARNOLD, B. (1985), p-Norm bounds on the expectation of the maximum of a possibly dependent sample, *Journal of Multivariate Analysis* **17**, 316-332

[2] BECK, A. (1964), On the linear search problem, *Israel J. Math.* **2**, 221-228.

[3] DAVID, H.A. (1981), *Order Statistics*, 2nd ed., Wiley, New York.

[4] ELTON, J. AND T. HILL (1992), Fusions of a probability distribution, *Ann. Probab.* **20**, no. 1, 421-454.

[5] Foley, R., T. Hill, and M. Spruill (1990), A generalization of Lévy's concentration-variance inequality, *Probability Theory and Related Fields* **86**, 53-62.

[6] Hartley, H.O. and H.A. David (1954), Universal bounds for mean range and extreme observation, *Annals of Mathematical Statistics* **25**, 85-99.

[7] Hengartner, W. and R. Theodorescu (1973), *Concentration functions*, Academic Press, New York.

[8] Lévy, P. (1937), *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris.