P8A.5

THE LARGEST PARTICLE IN A SAMPLE FROM AN EXPONENTIAL, GAMMA OR LOGNORMAL DISTRIBUTION

Paul L. Smith*
South Dakota School of Mines and Technology, Rapid City, SD USA

ABSTRACT

The size of the largest particle in a sample of hailstones or raindrops is of some interest. In the case of hail, it provides an indication of the potential for damage. In the case of rain, some scientists take it as an indication of the need to truncate any applicable drop-size distribution function. Such samples come from much larger populations of hydrometeors, and a simple *gedanken* experiment shows that the largest particle in a sample is most unlikely to be the largest particle in the population. This paper examines the sampling distribution of the maximum particle size, as a function of the sample size (number of particles observed).

This distribution is calculated analytically for exponential size distributions, and numerical results are provided for gamma and lognormal distributions. The results show that the maximum particle size in a sample is unlikely to approach the maximum size in the underlying population, even with a sample of many thousands of particles. Consequently, an observation of "maximum particle size" provides little more than a lower bound on the maximum size in the population, and establishes no basis for truncating any applicable size distribution function.

1. INTRODUCTION

The size of the largest raindrop, or the largest hailstone, in a population of precipitation particles is of interest. Mathematical functions used to describe the population size distribution could, and perhaps should, be truncated at the largest size (e.g., Ulbrich 1983, 1985; Ulbrich and Atlas 1998). In the case of hailstones, the maximum size can be a useful indicator of damage potential (e.g., Morgan 1982) and also provides clues regarding the hailstone growth process. However, the largest particle found in a sample taken from the population is not a reliable indicator of the size of the largest one in the underlying population – an issue treated using empirical data for hailstones in Smith and Waldvogel (1989) and using Monte Carlo simulations in Smith *et al.* (1993) and Smith and Kliche (2005).

A simple *gedanken* experiment shows that the largest precipitation particle in a sample is unlikely to represent the size of the largest one in the population being sampled. Suppose two size-measuring instruments, such as raindrop disdrometers or hailpads, were placed near enough, say a couple of meters apart, that they could be considered to provide independent samples from the same population. Large hydrometeors are sparsely distributed, so rarely if ever would the largest particles in concurrent samples from the two instruments have the same size (assuming the instrument size resolution provides adequate discrimination). Consequently, there would be at most a 50% chance that the largest particle in one of the samples represents the largest size in the population. Now add a third sensor, and so on; the probability that any one sample contains the largest particle would be inversely proportional to the number of instruments, and would become vanishingly small as that number grows.

Knowledge of the relationship between the largest size in a sample and the largest size in the underlying population would be of interest. Only the inverse of this relationship is readily obtainable, in the form of the probability distribution of the maximum particle size expected in a sample of any specified size taken from a specified population. If the probability density function (PDF) of particle sizes in the underlying population is known, the PDF of maximum size in a sample from the population can be calculated for any given sample size (e.g., Rice 1995). For the size distribution functions commonly used to represent precipitation particles (exponential, gamma, lognormal), this distribution can be readily calculated. Such calculations, summarized herein, illustrate the implications of the foregoing *gedanken* experiment.

2. GENERAL TREATMENT

Consider a population of spherical precipitation particles in the atmosphere described by some particle size distribution function (PSD) $n(D)$, where D represents the particle diameter. We write the PSD as a product of the total number concentration N_T and the probability density function of particle size $p(D)$; thus

$$n(D) = N_T\, p(D) \tag{1}$$

For present purposes we allow the PDF to extend to infinite diameter, clearly an unrealistic assumption. Truncation at some physically plausible maximum diameter would add a minor complication to the calculations. However, as will be shown below, such truncation is often irrelevant for typical sample sizes.

**Corresponding author address:* Paul L. Smith, Institute of Atmospheric Sciences, SDSM&T, 501 East Saint Joseph Street, Rapid City, SD 57701; e-mail: Paul.Smith@sdsmt.edu.

The probability that any given particle in the population is smaller than some specified threshold size t is

$$P(D \le t) = \int_0^t p(D)\,dD \qquad (2)$$

For a sample of C particles taken at random from the population, the probability that the largest particle D_{max} is no larger than t is

$$P(D_{\max} \le t) = [P(D \le t)]^C \qquad (3)$$

From this expression we can calculate the PDF of D_{max} as

$$p(D_{\max} = t) = \frac{d}{dt}[P(D_{\max} \le t)] \qquad (4)$$

For exponential PSDs the results are available in closed form, while for gamma or lognormal PSDs numerical techniques must be used.

3. RESULTS

a. Exponential Particle Size Distributions

For an exponential PSD,

$$n(D) = n_0 \exp(-\Lambda D) = N_T \Lambda \exp(-\Lambda D) \qquad (5)$$

where n_o and Λ are concentration ("intercept") and size ("slope") parameters, respectively. Thus

$$p(D) = \Lambda \exp(-\Lambda D) \qquad (6)$$

It is convenient to convert to a dimensionless particle size scale y by normalizing the diameters with respect to the mass-weighted mean diameter D_m: $y = D/D_m$. For an exponential PSD, $D_m = 4/\Lambda$, so the PDF can be written

$$p(D) = (4/D_m)\exp(-4D/D_m) \qquad (7)$$

$$p(y) = 4\exp(-4y) \qquad (8)$$

Let t_N represent the normalized threshold t/D_m, and $y_{max} = D_{max}/D_m$; then

$$P(y \le t_N) = \int_0^{t_N} p(y)\,dy = 1 - \exp(-4t_N) \qquad (9)$$

and in a sample of C particles from the exponential population,

$$P(y_{\max} \le t_N) = [1 - \exp(-4t_N)]^C \qquad (10)$$

The PDF of y_{max} then becomes

$$p(y_{\max} = t_N) = 4C\, e^{-4t_N} \left[1 - e^{-4t_N}\right]^{C-1} \qquad (11)$$

Figure 1 shows this PDF for three different sample sizes. The shape of the PDF varies only slightly for such large sample sizes, and the mode shifts to the right by (ln10)/4 for each factor 10 increase in sample size. (This shift can be seen in Fig. 3 of Smith and Waldvogel 1989.) While the PDF of particle size in (6) is not truncated at a maximum size, it is evident from Fig. 1 that even in a sample of 10,000 particles the probability of finding one as large as $D = 3\ D_m$ is small. For example, with a raindrop population having $D_m = 2$ mm this means that a drop as large as 6 mm is unlikely to appear in such a sample. That is a plausible raindrop size, so such a sample would provide no basis for truncating the PSD at D_{max}.

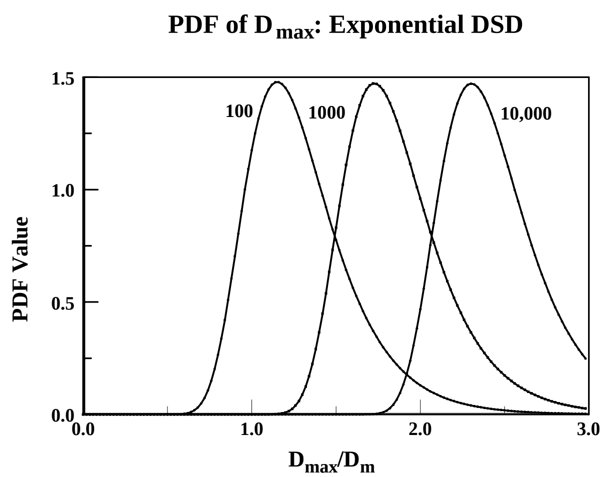

Fig. 1: Probability density functions for the (normalized) maximum particle size in samples from an exponential size distribution, for sample sizes as indicated on the curves.

To be sure, if the sample size exceeded 100,000 raindrops the mode of the PDF would approach $D_{max}/D_m = 3.0$ and the chance of finding a larger drop would become appreciable. But such large samples could only be obtained by commingling data from different times and/or places, raising issues about homogeneity of the samples.

It is also true that sample values of raindrop D_m can be larger, perhaps up to 3 mm; in such a case $y_{max} = 3.0$ means $D_{max} = 9$ mm, an implausibly large size for a raindrop. However, such occurrences are not common; moreover, considering the uncertainties in the sample estimates of D_m (Smith *et al.* 1993; Smith and Kliche 2005) the largest values are likely to be overestimates.

Figure 2 presents the results in a different form, showing the probability that the largest value of y = D/D_m in a sample is no greater than that indicated by the values on the curves, as a function of sample size. For instance, in samples of 2000 particles the median value of D_{max} (indicated by the 0.5 probability point on the ordinate) is about 2.0 D_m. Even in samples of 100,000 particles, the median D_{max} is still less than 3.0 D_m.

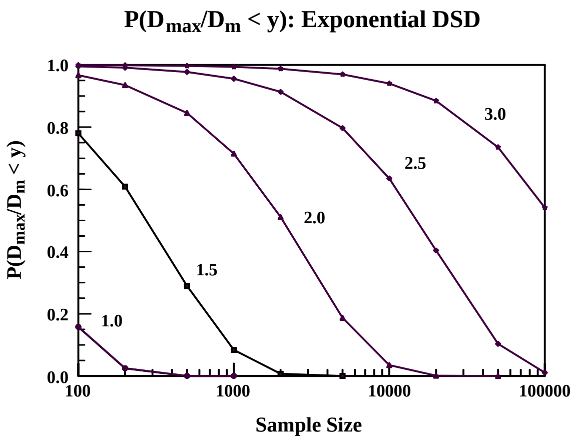

Fig. 2: Probability that the (normalized) size of the largest particle in a sample of the size indicated on the abscissa, taken from an exponential distribution, is no greater than the value of y indicated on the respective curves.

b. Gamma Particle Size Distributions

For a gamma PSD the PDF of particle size, equivalent to (6), is

$$p(D) = \frac{\lambda^{\mu+1}}{\Gamma(\mu+1)} D^{\mu} \exp(-\lambda D) \qquad (12)$$

where λ is the scale (size) parameter, μ is the distribution shape parameter, and $\Gamma(x)$ denotes the gamma function. Here $D_m = (\mu + 4)/\lambda$; thus

$$p(D) = \frac{(\mu+4)^{\mu+1}}{\Gamma(\mu+1)} \frac{D^{\mu}}{D_m^{\mu+1}} e^{-(\mu+4)D/D_m} \qquad (13)$$

Converting to normalized diameters,

$$p(y) = \frac{(\mu+4)^{\mu+1}}{\Gamma(\mu+1)} y^{\mu} e^{-(\mu+4)y} \qquad (14)$$

The cumulants called for in (2) – (4) are incomplete gamma functions that cannot be expressed in closed form, requiring numerical computations. If P_g represents the cumulant corresponding to (2), then

$$P(y_{\max} \le t_N) = (P_g)^C \qquad (15)$$

and

$$p(y_{\max} = t_N) = C \frac{(\mu+4)^{\mu+1}}{\Gamma(\mu+1)} t_N^{\mu} e^{-(\mu+4)t_N} (P_g)^{C-1} \qquad (16)$$

Figure 3 shows the PDF of D_{max} for the case of shape parameter μ = 2, and three different sample sizes. Here the PDF narrows, and the tendency of the mode to shift to the right diminishes, as the sample size increases. There would be very little chance of finding a particle approaching D = 3.0 D_m in samples of plausible sizes from this PSD. Gamma size distributions with higher values of the shape parameter are even narrower, reducing the probability of finding a particle that large even further.

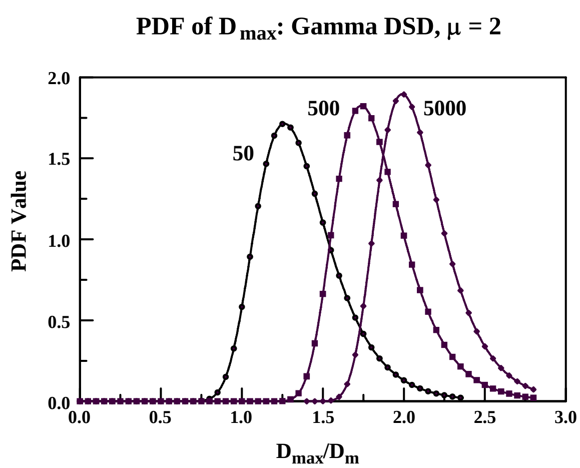

Fig. 3: Probability density functions for the (normalized) maximum particle size in samples from a gamma size distribution with shape parameter μ = 2, for sample sizes as indicated on the curves.

Figure 4 again presents the information in a format comparable to Fig. 2. The curves for y = 2.5 in Figs. 2 and 4 are nearly identical, but the probability that the largest particle is smaller than 3.0 D_m in the gamma distribution of Fig. 4 is even greater than is the case for the exponential PSD of Fig. 2. The exponential PSD is a limiting case of the gamma family, with shape parameter μ = 0, and is the one with the heaviest tail.

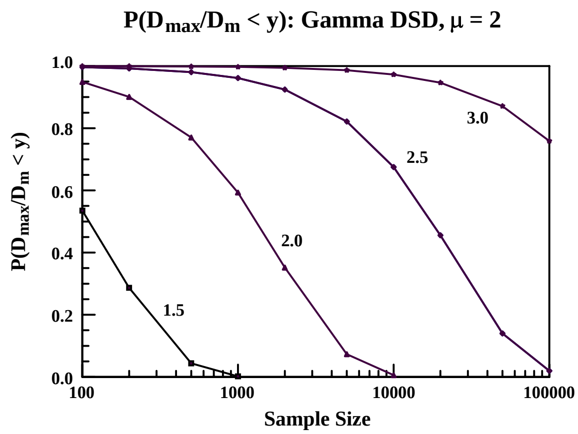

Fig. 4: Probability that the (normalized) size of the largest particle in a sample of the size indicated on the abscissa, taken from a gamma distribution with μ = 2, is no greater than the value of y indicated on the respective curves.

c. Lognormal Particle Size Distributions

For a lognormal PSD the PDF of particle size, equivalent to (6), is (Feingold and Levin 1986)

$$p(D) = \frac{1}{D(\ln\sigma)\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(D/D_g)}{\ln\sigma}\right]^2\right\} \quad (17)$$

where $\ln\sigma$ is the distribution shape parameter and D_g is the scale (size) parameter.

In terms of normalized diameters (17) becomes

$$p(y) = \frac{1}{y(\ln\sigma)\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(y/y_g)}{\ln\sigma}\right]^2\right\} \quad (18)$$

where y_g is a normalized scale parameter $y_g = D_g/D_m$. For a lognormal PSD, the definition of D_m as the ratio of the fourth to the third moment of the PSD yields $D_m = D_g \exp[7(\ln\sigma)^2/2]$; therefore

$$y_g = \exp[-7(\ln\sigma)^2/2] \quad (19)$$

The cumulant for this PDF is

$$P(y \le t_N) = \frac{1}{2}\left\{1 + \mathrm{erf}\left[\frac{\ln(t_N/y_g)}{\sqrt{2}(\ln\sigma)}\right]\right\} \quad (20)$$

The cumulants called for in (2) – (4) now involve “error functions” erf(x) that also cannot be expressed in closed form, again requiring numerical computations.

Once again the equivalent of (15) can be used to calculate the probability that no particle in a sample of specified size exceeds the threshold size t, as in (3); the derivative of that result would yield the PDF of D_{max}. In this case, however, the latter calculations are substantially more complicated. Thus here we only use (20) with (3) to calculate the values for Fig. 5, which corresponds to the preceding Figs. 2 and 4. Comparison of Fig. 5 with Fig. 2 shows that the probability that the largest particle could exceed 3.0 D_m in the specified lognormal distribution is greater than is the case for the exponential PSD, with comparable sample sizes. The wider lognormal distributions have heavy tails, and in that case truncation of the PDF of particle size may be essential. Nevertheless, the curves of Fig. 5 demonstrate that the size of the largest particle in a sample is generally a poor indicator of the appropriate truncation point.

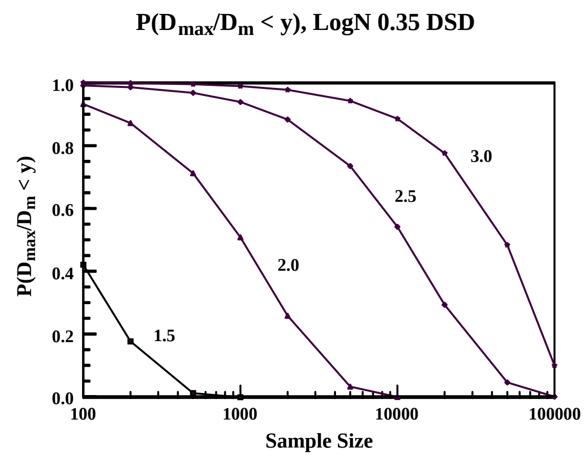

Fig. 5: Probability that the (normalized) size of the largest particle in a sample of the size indicated on the abscissa, taken from a lognormal distribution with shape parameter $\ln\sigma$ = 0.35, is no greater than the value of y indicated on the respective curves.

4. CONCLUSIONS

The maximum particle size in a sample of raindrops or hailstones is unlikely to approach the maximum size in the underlying population, even with a sample of many thousands of particles. For example, in a sample of 5,000 drops from an exponential distribution with mass-weighted mean diameter 2 mm (D_o = 1.84 mm) there would be only a 20% chance of finding a drop as large as 5 mm, even if the distribution extends to infinite size. Consequently, an observation of “maximum particle size” provides little more than a lower bound on the maximum size in the population. It clearly establishes no basis for truncating any applicable size distribution function.

Acknowledgment: The State of South Dakota provided partial support for preparation of this paper.

5. REFERENCES

Feingold, G., and Z. Levin, 1986: The lognormal fit to raindrop spectra from frontal convective clouds in Israel. *J. Climate Appl. Meteor.*, **25**, 1346-1363.

Morgan, G.M., Jr., 1982: Observation and measurement of hailfall. In *Thunderstorms: A Social, Scientific and Technological Documentary, Vol. 3, Instruments and Techniques for Thunderstorm Observations and Analysis.* E. Kessler (ed.), U.S. Dept. of Commerce, 149-180.

Rice, J.A., 1995: *Mathematical Statistics and Data Analysis* (2nd Ed.). Duxbury Press, pp. 108-109.

Smith, P. L., Jr. and A. Waldvogel, 1989: On determinations of maximum hailstone sizes from hailpad observations. *J. Appl. Meteor.*, **28,** 71-76.

Smith, P.L, Z. Liu, and J. Joss, 1993: A study of sampling variability effects in raindrop size observations. *J. Appl. Meteor.*, **32**, 1259-1269.

Smith, P. L. and D. V. Kliche, 2005: The bias in moment estimators for parameters of drop-size distribution functions: sampling from exponential distributions. *J. Appl. Meteor.*, **44,** 1195-1205.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distribution. *J. Climate Appl. Meteor.*, **22**, 1764-1775.

Ulbrich, C. W., 1985: The effects of drop size distribution truncation on rainfall integral parameters and empirical relations. *J. Climate Appl. Meteor.,* **24**, 580-590.

Ulbrich, C.W., and D. Atlas, 1998: Rainfall microphysics and radar properties: analysis methods for drop size spectra. *J. Appl. Meteor.*, **37**, 912-923.