Today, we will:

- Discuss **particle statistics**: grouped data vs. listed data, histograms, PDFs, and cumulative distribution functions
- Analyze in detail a sample of polluted air from Hinds' air pollution book. See class handout, and also the posted Excel spreadsheet on the course website.

AEROSOL PARTICLE STATISTICS

Fundamentals & Definitions

$\boxed{\text{Sample}}$ = a collection of particles to be studied

$\boxed{\text{Listed data}}$ → We are used to in statistics → each data pt. is listed with its own properties

eg

[Typically plot histograms with Δx = constant]

data pt		$x = D_p$ (μm)
i = 1	j = 1	0.5
2		0.7
3		0.75
4	j = 2	1.1
5		1.1
6		1.4
7		1.9
8	j = 3	2.1
:		:

(increasing order)

Can draw histogram, create PDFs, calc. statistics

In air pollution we typically have $\boxed{\text{grouped data}}$ → properties known only in bins or groups or classes

We don't know properties of individual particles

For same e.g. as above:

e.g. ~~Cascade impactor~~

Δx is not constant

j = class # or bin #	D_p range (μm)	(middle value) D_j (μm)	mass of whole bin M_j (mg)	# particles in bin j (if known) n_j
1	0.1 – 1	0.55	6.2×10^{-6}	3
2	1 – 2	1.5	1.8×10^{-5}	4
3	2 – 4	3.0	2.9×10^{-5}	:
:	$D_{p,min,j}$ 4 – 8 $D_{p,max,j}$	→ 6.0	1.3×10^{-4}	:
J				

let m_t = total mass = $\sum_{j=1}^{J} m_j$

n_t = total # particles = $\sum_{j=1}^{J} n_j$

Need to come up with methods to analyze grouped data statistically

Example: Converting from mass distribution to number distribution

Given: A cascade impactor is used to sample air in a city during a temperature inversion. Tray number 5 of the impactor weighs 13.20 mg clean. After the sample is taken, the same tray (with collected particles) weighs 13.32 mg. This tray (Tray 5) has been calibrated to collect particles between 4 and 6 microns. Assume unit density spherical particles (ρ_p = 1000 kg/m^3), and that the air is at STP, ρ = 1.184 kg/m^3, μ = 1.849 × 10^{-5} kg/(m s).

To do: Estimate the number of particles collected on Tray 5. *Hint*: We approximate *all* the particles on Tray 5 as having the *middle* diameter in its range, i.e., 5 microns (half-way between 4 and 6 microns). Give your answer in millions of particles to 3 significant digits.

Solution:

Comment: This is typical → we have mass data (grouped data), not number data.
So – we need to convert from mass distribution to number distribution for each bin or class j.

$$M_j = n_j \cdot (\text{mass per particle}) = (13.32 - 13.20)\text{ mg} = \underline{0.12 \text{ mg} = M_j}$$

$$M_j = n_j \rho_p \frac{\pi}{6} D_p^3$$

$$n_j = \frac{6 M_j}{\rho_p \pi D_p^3} = \frac{6(0.12\text{ mg})}{\left(1000 \frac{\text{kg}}{\text{m}^3}\right)\pi\left(5.0\times10^{-6}\text{ m}\right)^3}\left(\frac{1\text{ kg}}{10^6\text{ mg}}\right)$$

$$= 1.833 \times 10^6 \text{ particles}$$

$$\boxed{1.83 \text{ million particles}}$$

The data shown below are from the class handout, so you can follow along.

j = class (bin number)	$D_{p,\min,j}$ (lower limit) (μm)	$D_{p,\max,j}$ (upper limit)	$D_{p,j}$ (middle value)	$\Delta D_{p,j}$ = class width	n_j = frequency (count per class)	$n_j/\Delta D_{p,j}$ = count per class width	$f(D_{p,j})$ = $n_j/(\Delta D_{p,j} n_t)$ = fraction per class width	probability in this class = $f(D_{p,j})*\Delta D_{p,j}$ = n_j/n_t
1	1	4	2.5	3	104	34.667	0.0347	0.104
2	4	6	5	2	160	80.000	0.0800	0.16
3	6	8	7	2	161	80.500	0.0805	0.161
4	8	9	8.5	1	75	75.000	0.0750	0.075
5	9	10	9.5	1	67	67.000	0.0670	0.067
6	10	14	12	4	186	46.500	0.0465	0.186
7	14	16	15	2	61	30.500	0.0305	0.061
8	16	20	18	4	79	19.750	0.0198	0.079
9	20	35	27.5	15	90	6.000	0.0060	0.09
10	35	50	42.5	15	17	1.133	0.0011	0.017
11	50	100	75	50	0	0.000	0.0000	0

Minimum diameter in class j

Maximum diameter in class j

Middle diameter in class j: $D_{p,j} = (D_{p,\min,j} + D_{p,\max,j}) / 2$

Width of class j: $\Delta D_{p,j} = D_{p,\max,j} - D_{p,\min,j}$

Number of particles in class j (see Plot 1)

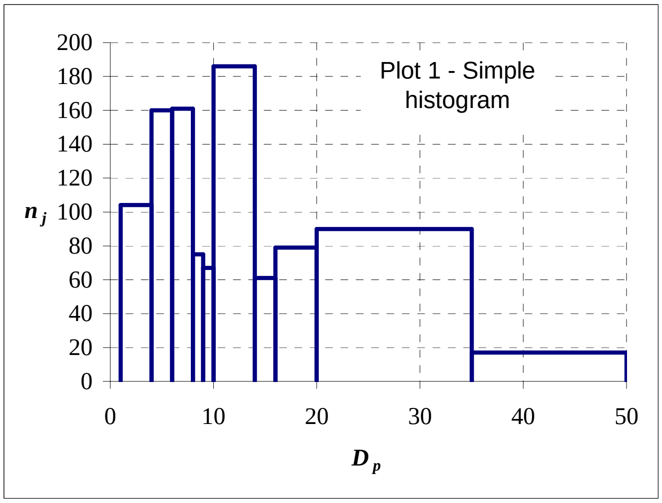
Number of particles in class j divided by class width: $n_j/\Delta D_{p,j}$ (see Plot 2)

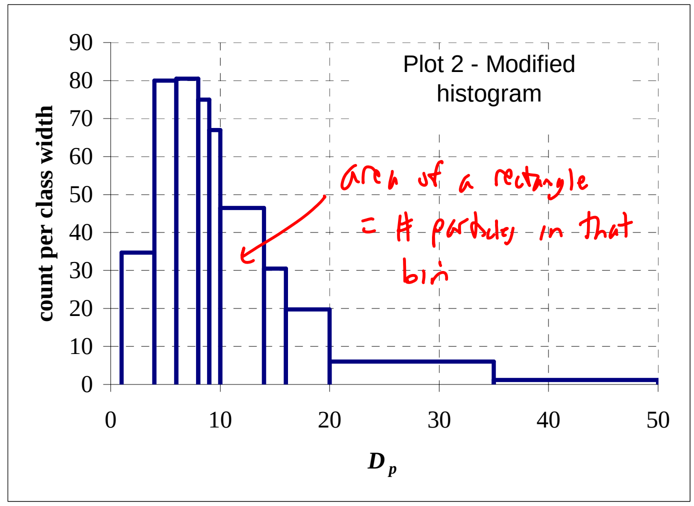
Fraction of particles in class j divided by class width: $n_j/(\Delta D_{p,j} n_t)$ (see Plot 3)

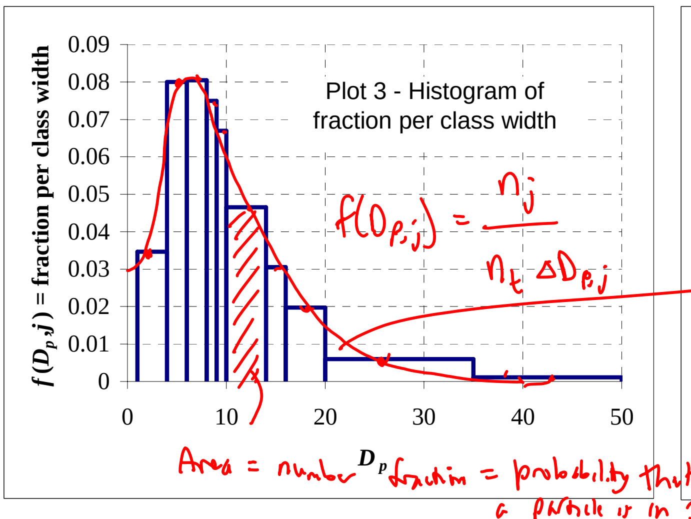
Probability in class j: $f(D_{p,j})*\Delta D_{p,j}$ also = n_j/n_t

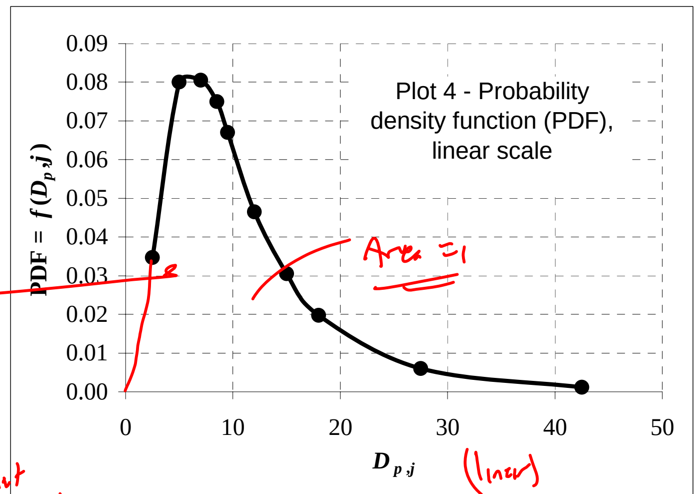
add these one at a time to generate the Cumulative distribution function $F(D_p)$

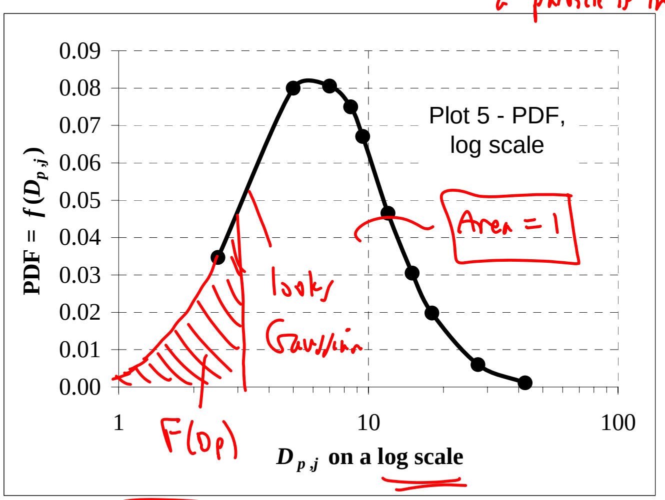
Probability $\frac{n_j}{n_t}$	$F(D_p)$
0.104	0.104
0.16	0.104 + 0.16 = 0.264
0.161	0.264 + 0.161 = 0.425
⋮	etc. – $\boxed{\text{SEE EXCEL FILE}}$ ☆

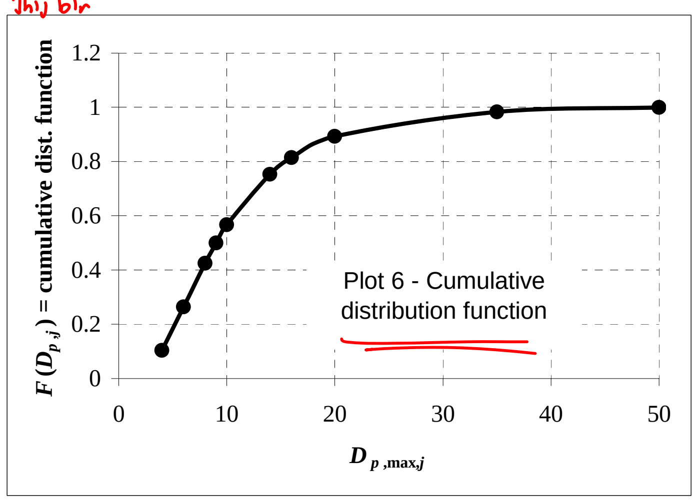






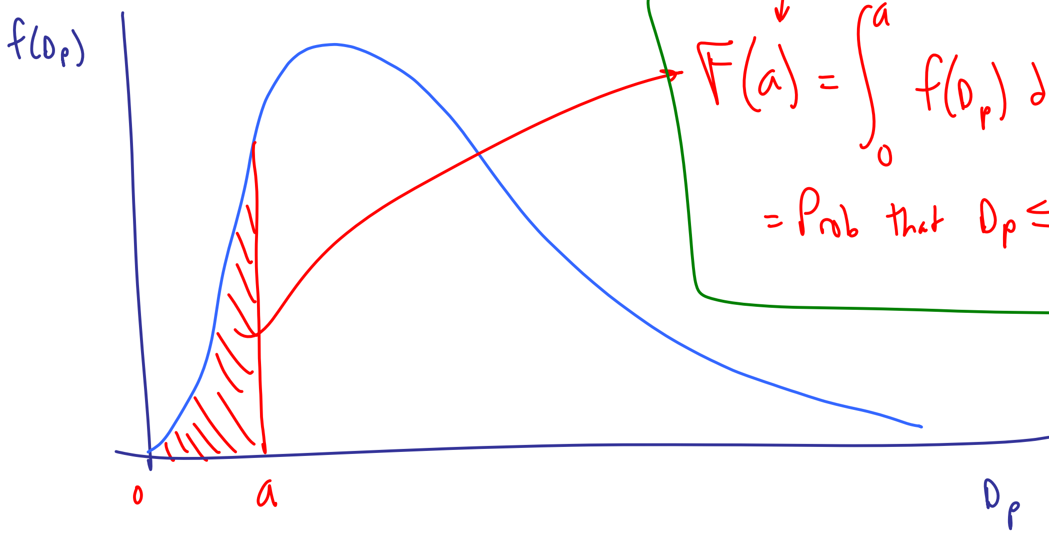






$$\boxed{\int_0^\infty f(D_p)\, dD_p = 1} \rightarrow \text{For any PDF}$$

Cumulative Distribution Function



$D_p = a$

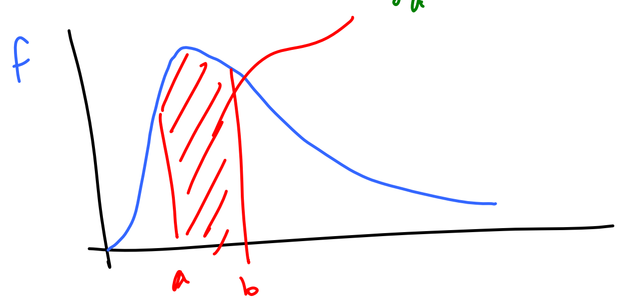
$$F(a) = \int_0^a f(D_p)\, dD_p$$

$= \text{Prob that } D_p \leq a$

Comments: $F(\infty) = 1$ (all particles) $\rightarrow$ 100%

$$F(b) - F(a) = \int_0^b f(D_p)\, dD_p - \int_0^a f(D_p)\, dD_p$$

$$= \int_a^b f(D_p)\, dD_p$$



Special cases $\rightarrow$ At one value of D_p, $F(D_p) = 0.5 \rightarrow 50\%$

$\boxed{\text{At } D_{p,50} = \text{median diameter}}$ $\rightarrow$ $D_{p,50}$ (number)

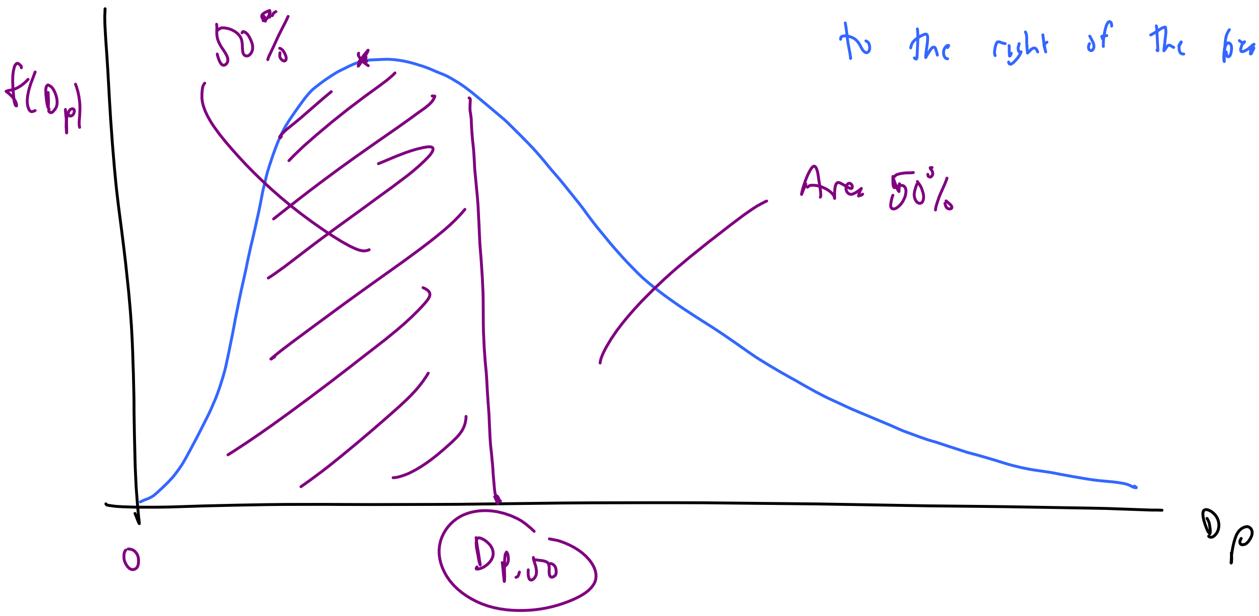
based on # particles

We will also define $D_{p,50}$ (mass) based on mass

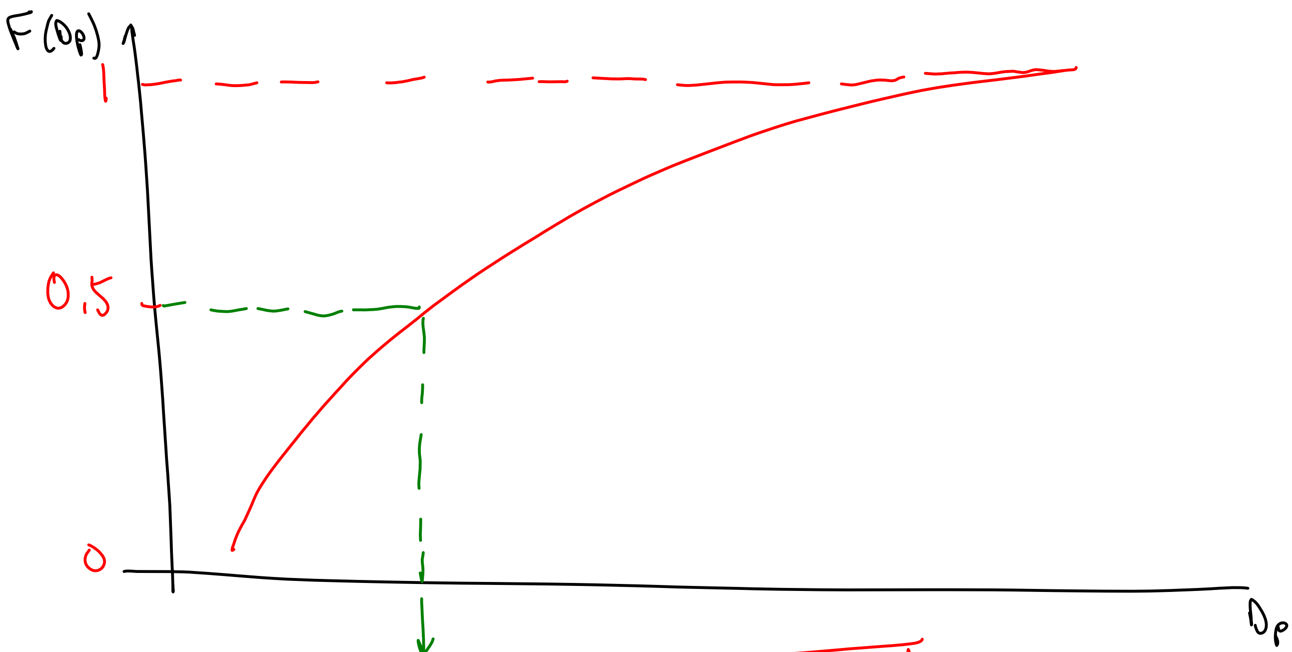
$$\boxed{F(D_{p,50}) = 0.50}$$

$$\boxed{D_{p,50}\text{ (mass)} \neq D_{p,50}\text{ (number)}}$$

FOR A LOGNORMAL PDF _ $D_{p,50}$ (number) occurs to the right of the peak



Finally, Plot 6 of the handout is $F(D_p)$ for D_p from smallest to largest



$D_{p,50}$ = median particle diameter = where $F(D_p) = 50\%$