**Today, we will**:

"GROUPED DATA"

- Discuss **Particle Statistics** in **Section 8.2**
- Discuss an example particle size distribution in detail – ***see class handout***

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

| $j$ = class (bin number) | $D_{p,\min,j}$ (lower limit) | $D_{p,\max,j}$ (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 | areas → | 0.104 |
| 2 | 4 | 6 | 5 | 2 | | 160 | 80.000 | 0.0800 | in → | 0.16 |
| 3 | 6 | 8 | 7 | 2 | not | 161 | 80.500 | 0.0805 | plot 3 | 0.161 |
| 4 | 8 | 9 | 8.5 | 1 | constant | 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 |

total = 1000

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

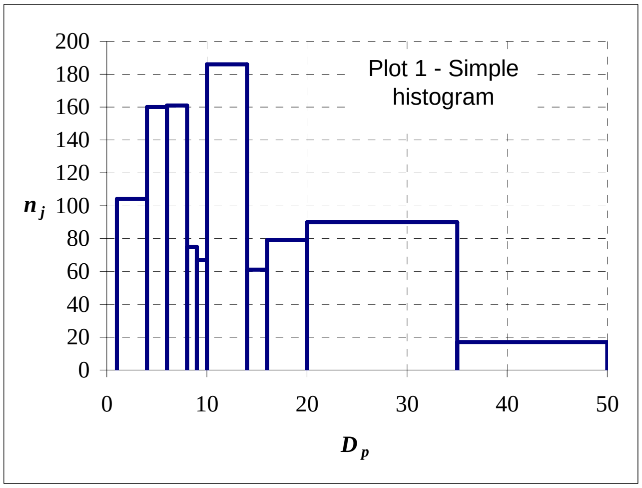
Number of particles in class $j$ (see Plot 1)

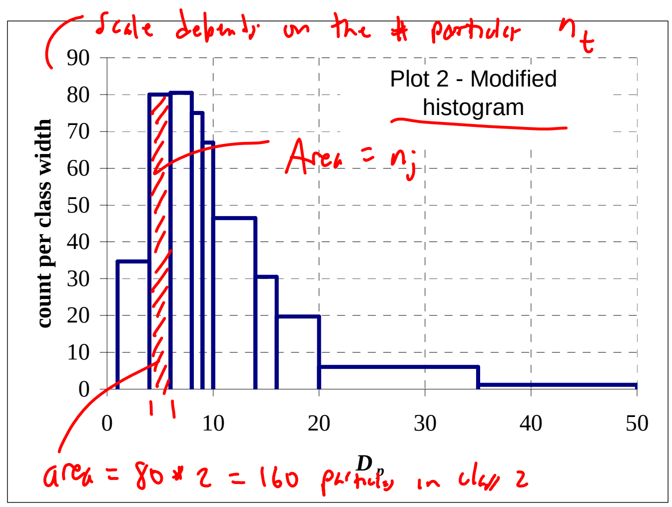
Number of particles in class $j$ divided by class width: $n_j/\Delta D_{p,j}$ (see Plot 2)

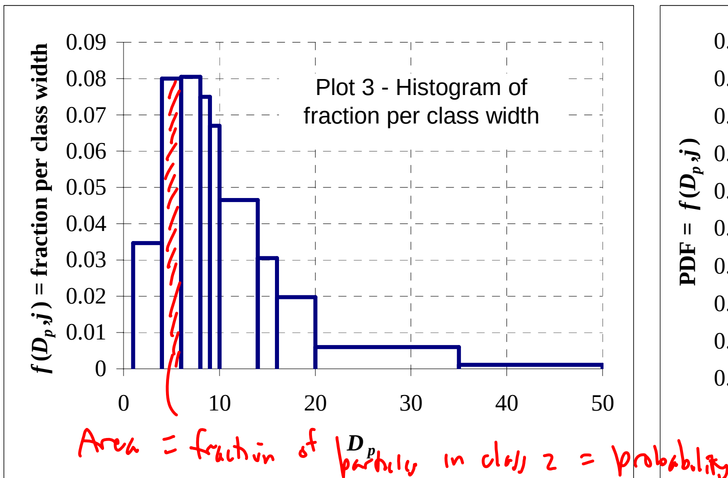
Fraction of particles in class $j$ divided by class width: $n_j/(\Delta D_{p,j} n_t)$ (see Plot 3)

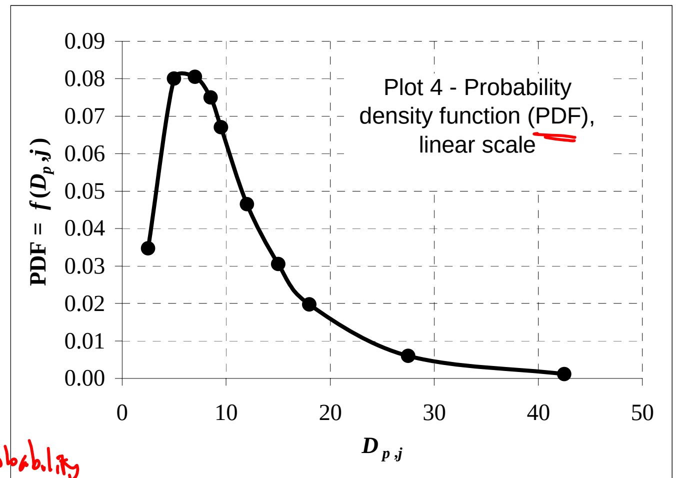
Probability in class $j$

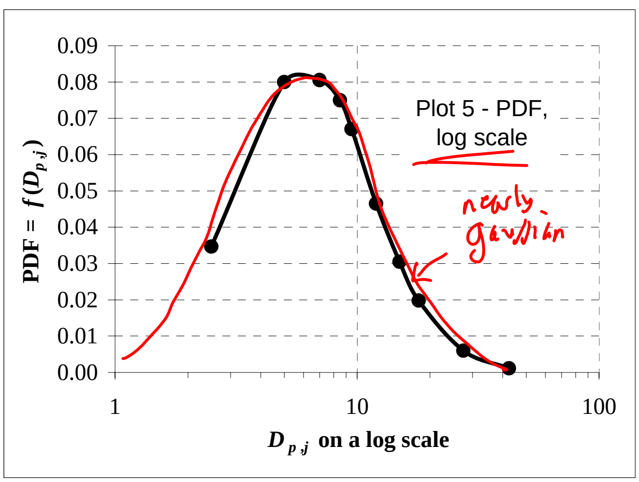
See the handout – also available as an Excel file on the website.

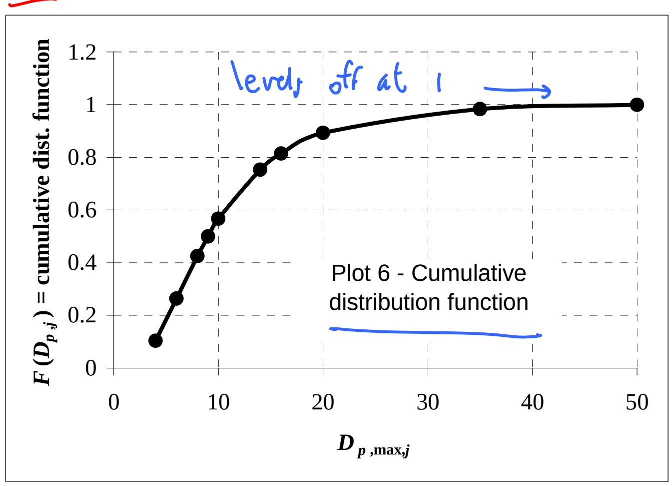


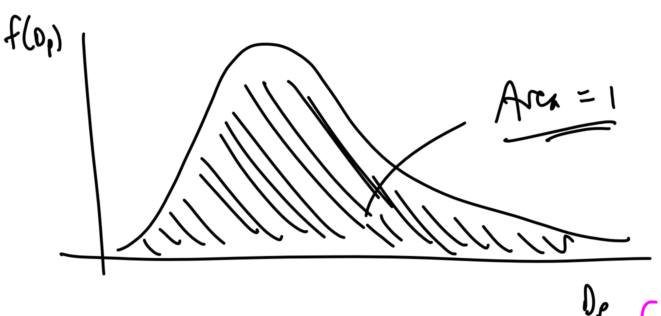




Plot 4 = PDF




$$\boxed{\int_0^\infty f(D_p)\,dD_p = 1}$$

(This is true for any PDF)

Our PDF is not symmetric — Not gaussian

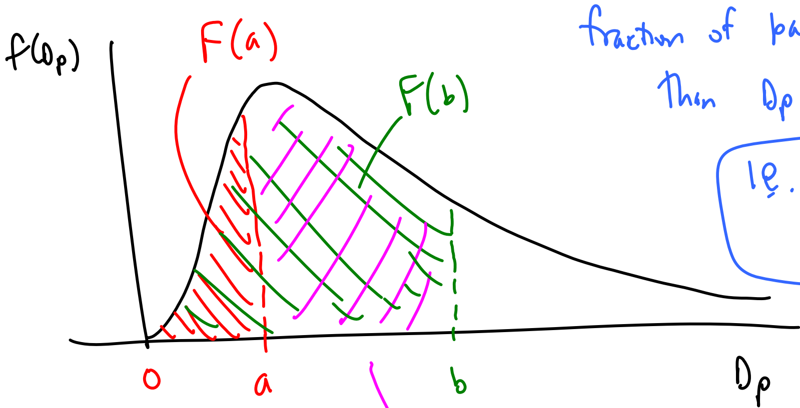
However, it turns out that many particle distributions in air pollution work are log-normal — see plot 5

## Cumulative Distribution Function $F(D_p)$

$$\boxed{\text{For } D_p = a, \quad F(a) = \int_0^a f(D_p)\, dD_p}$$ by definition

= fraction of particles that are smaller than $D_p = a$

i.e. – The probability that $D_p < a$




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

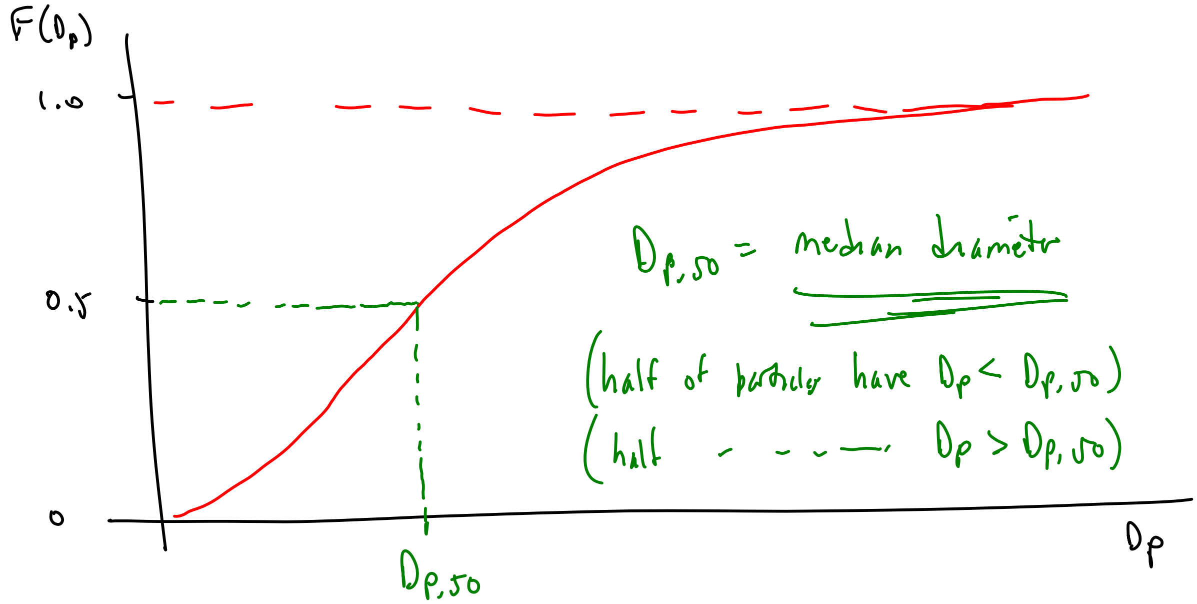
= Prob. that a particle diameter lies between a & b

$F(\infty) = 1$ (Prob. that $0 < D_p < \infty$ is 100%!)

$$F(a) - F(a) = \int_a^a f(D_p)\, dD_p = 0$$

Zero prob. that $D_p = a$ exactly

Cumulative Distrib. Function:




$D_{p,50}$ = median diameter

(half of particles have $D_p < D_{p,50}$)
(half - - - - $D_p > D_{p,50}$)

---

For our example

$D_{p,50} = D_{p\,max}$ at which $F(D_p) = 0.500$

→ here, class $j=4$

$\boxed{D_{p,50} = 9.0\ \mu m}$  [Not 8.5 μm ($D_{p,j}$) middle value]

max value in the class $j=4$ ✩

50% of the particles in this sample are < 9.0 μm in diameter
50% - - - - - - - > - - - - - ✩

---

STATISTICS: Arithmetic mean diameter = $D_{p,am}$ (number)

based on number of particles

[later on we will define $D_{p,am}$ (mass)]

For listed data

$$\boxed{D_{p,am} = \frac{\sum_{i=1}^{n_t} D_{p,i}}{n_t}}$$

For grouped data: Assume each particle in a particular bin has $D_p = D_{p,j}$ (mid value)

$$\boxed{D_{p,am} = \sum_{j=1}^{J} \frac{n_j D_{p,j}}{n_t}} \quad ☆$$

Here, we get $\boxed{D_{p,am} = 11.2\ \mu m}$ ☆

NOTICE: $D_{p,am} \neq D_{p,median}$ (since the PDF is not Gaussian)

---

Standard Deviation

For listed data,

$$\boxed{\sigma = \left[\frac{\sum_{i=1}^{n_t} \left(D_{p,i} - D_{p,am}\right)^2}{n_t - 1}\right]^{1/2}}$$

For grouped data,

$$☆ \quad \boxed{\sigma = \left[\frac{\sum_{j=1}^{J} n_j \left(D_{p,j} - D_{p,am}\right)^2}{n_t - 1}\right]^{1/2}}$$

Here, we calculate $\underline{\sigma = 7.93\ \mu m}$

---

$D_{p,mode}$ = highest point on the PDF (the most frequent class)

$$\boxed{D_{p,mode} \simeq 7.0\ \mu m}$$

Here, $\boxed{D_{p,mode} < D_{p,50} < D_{p,am}}$

· Geometric mean $\quad D_{p,gm}$ $\quad$ [use multiplication instead of addition]

listed data:

$$\boxed{D_{p,gm} \equiv \left( \underset{\text{particle 1}}{D_{p_1}} \; \underset{\text{particle 2}}{D_{p_2}} \; D_{p_3} \cdots\cdots D_{p_{n_t-1}} \; D_{p_{n_t}} \right)^{\frac{1}{n_t}}}$$

For grouped data,

$$\boxed{D_{p,gm} = \left( \underset{\text{class \#1, } j=1}{D_{p,1}^{\,n_1}} \; D_{p,2}^{\,n_2} \; D_{p_3}^{\,n_3} \cdots D_{p_J}^{\,n_J} \right)^{\frac{1}{n_t}}}$$

☆

OR, after some algebra

$$\boxed{D_{p,gm} = \exp\left[ \frac{1}{n_t} \sum_{j=1}^{J} n_j \ln D_{p,j} \right]}$$

Here, $\boxed{D_{p,gm} = 8.98\ \mu m}$

---

Geometric standard deviation $\quad \sigma_g$

Define $\sigma_g$ using its natural log.

$\boxed{\ln \sigma_g \equiv \text{standard deviation in terms of } \ln(D_{p,j}) \text{ instead of } D_{p,j} \text{ itself.}}$

$$\boxed{\sigma_g = e^{\ln \sigma_g}}$$

$$\boxed{\ln \sigma_g = \left[ \frac{\sum_{j=1}^{J} n_j \left[ \ln(D_{p,j}) - [\ln(D_{p,j})]_{am} \right]^2}{n_t - 1} \right]^{1/2}}$$

For our sample data, $\boxed{\sigma_g = 1.97\ \mu m}$

where $[\ln(D_{p,j})]_{am} = \boxed{[\ln(D_p)]_{am} = \frac{\sum_{j=1}^{J} n_j \ln(D_{p,j})}{n_t}}$