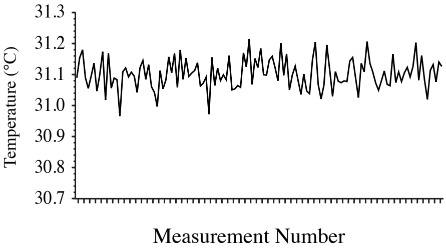
Chapter 02

Probability Distributions

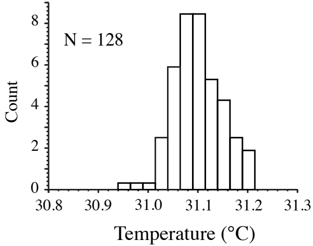
P. J. Grandinetti

Chem. 4300

Temperature of solution in "constant temperature" bath



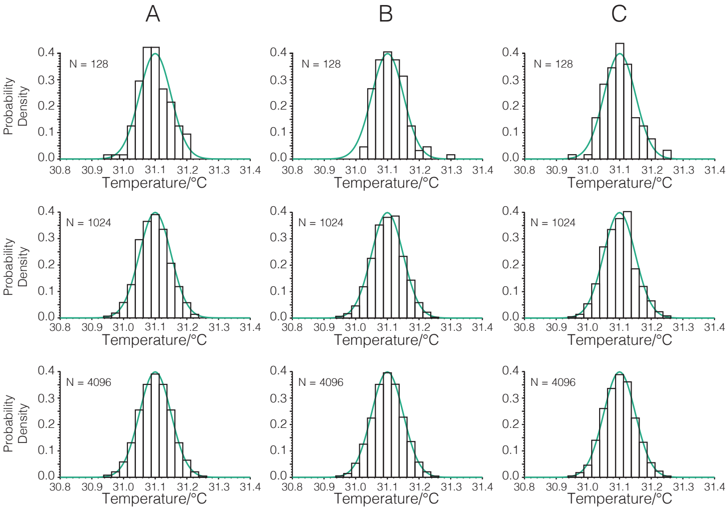
histogram,
aka *the sample distribution*



What do you report for the solution temperature?

- the mean temperature (31.1 °C)
- histogram of all measured values

Sample and parent distributions



- In $N \to \infty$ limit, the histogram or *sample distribution* becomes the *parent distribution*.
- Histogram of measured values is governed by an underlying parent probability distribution.

Probability density

Our main objective in making a measurement is to learn the underlying parent distribution, $p(x)$, that predicts the spread in the measured values.

The parent distribution, $p(x)$, is also called a *probability density function*.

A probability density function is

1. positive for all values of x,
2. area under the distribution is unity,

$$\int_{\text{all } x} p(x)\, dx = 1$$

Normalizing a probability density function

Example

Can $f(x) = e^{-x^2}$ be a probability density function? If not, could $f(x)$ be scaled by a constant normalization factor to create a valid probability density function?

- $f(x)$ is positive for all x
- one can show—with help of integral tables—that integrated area under $f(x)$ over all x is not unity,

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}, \qquad \text{cannot serve as a probability density function since } \int_{\text{all } x} f(x)dx \neq 1 \;.$$

- Given that $f(x)$ is always positive it can be normalized using the result above to define

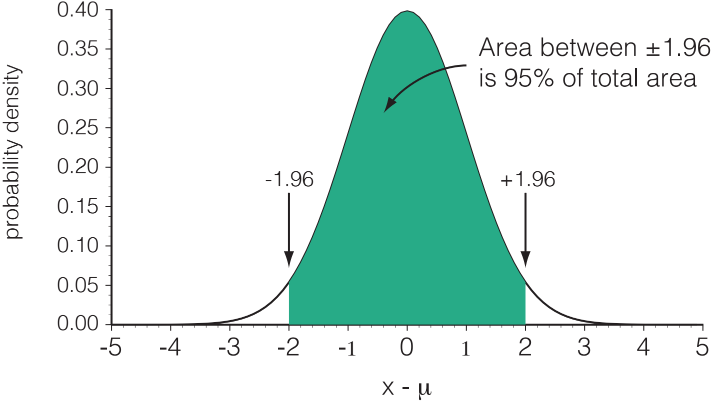
$$p(x) = \frac{1}{\sqrt{\pi}} f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$$

and then $p(x)$ can serve as a probability density function.

Confidence Limits

Probability that measured value lies between x_- and x_+ is calculated from parent distribution:

$$P(x_-, x_+) = \int_{x_-}^{x_+} p(x)\, dx$$



- Integral limits x_- and x_+ are called the *confidence limits* associated with probability $P(x_-, x_+)$.
- Note use of upper case P for probability and lower case p for probability density. A probability density is not a probability until is integrated between two limits.

The Mean

The mean describes the average value of the distribution

- From a series of measurements the true mean is given by

$$\mu = \lim_{N\to\infty} \frac{1}{N} \sum_i x_i$$

  where N corresponds to the number of measurements x_i.

- Practically, cannot make infinite measurements so estimate of mean, $\bar{x}$, is defined as

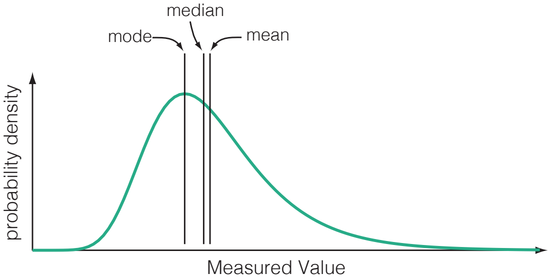
$$\bar{x} = \frac{1}{N} \sum_i x_i$$

- Given the parent distribution, $p(x)$, the mean is calculated according to

$$\mu = \int_{\text{all } x} x\, p(x)\, dx$$

The Mean, the Median, and the Mode.

When distribution is not symmetric about the mean ...



...two other parameters used are:

- median: cuts parent distribution area in half,

$$\int_{-\infty}^{x_{\text{median}}} p(x)\,dx = \frac{1}{2}.$$

- mode: most probable value,

$$\frac{dp(x_{\text{mode}})}{dx} = 0, \quad \text{and} \quad \frac{d^2p(x_{\text{mode}})}{dx^2} < 0.$$

The Variance.

Variance characterizes the width of the distribution

- From a series of measurements the true variance is obtained through:

$$\sigma^2 = \lim_{N\to\infty} \frac{1}{N} \sum_i^N (x_i - \mu)^2$$

- The estimate of the variance is defined as:

$$s^2 = \frac{1}{N-1} \sum_i^N (x_i - \overline{x})^2$$

- σ and s are the *standard deviation* are the true and estimated standard deviations.
- Given the parent distribution, $p(x)$, the variance is calculated according to

$$\sigma^2 = \int_{\text{all } x} (x-\mu)^2 p(x) dx$$