Physical Principles in Biology
Biology 3550
Fall 2017

Lecture 13

A Random Walk to Error Analysis

Wednesday, 20 September

©David P. Goldenberg
University of Utah
goldenberg@biology.utah.edu

Significant Figures

- From Quiz 1: Calculate the volume of a spherical yeast cell, with radius $5\,\mu\text{m}$

$$V = \frac{4}{3}\pi r^3$$

  - A correct and appropriate answer: $V = 500\,\mu\text{m}^3$
  - A problematic answer: $V = 524\,\mu\text{m}^3$
- The general rule: The final reported value should include no more significant figures than found in the input value with the smallest number of significant figures.
- Why this is important: Using more than the correct number of significant figures implies that the result is known with greater precision than it really is.
- Reporting extra significant figures is to misrepresent your results. It's a lie!

A Qualification to the Significant-figures Rule

- Using a calculated result as an input to another calculation:
- From Quiz 1: Calculate the number of Ca^2 ions in a yeast cell.

$$\text{moles} = \text{volume} \times \text{concentration}$$
$$= 5.24\,\mu\text{L} \times 10^{-7}\,\text{M}$$

- It's *good* to carry through "extra" significant figures in a calculation!
- This minimizes round-off error that would occur if the results were rounded at each step. Error grows with each round off.
- Round when the final result is reported.

A Rule For Rounding

- What to do when the digit to be rounded off is 5?
- An example, round to two digits:

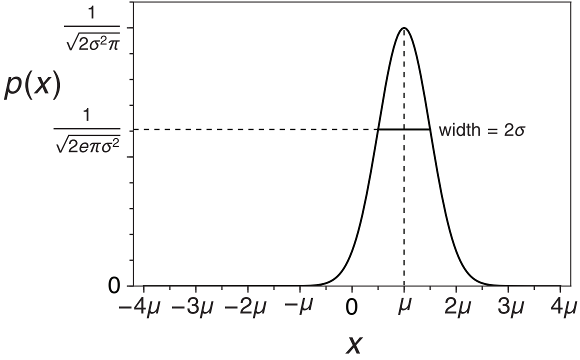
$$5.25 \longrightarrow 5.2$$

or

$$5.25 \longrightarrow 5.3$$

- Always round to the even digit.
- Prevents systematic rounding up or down.
- Always rounding to the odd digit would work, too! But, rounding to the even digit is the convention.

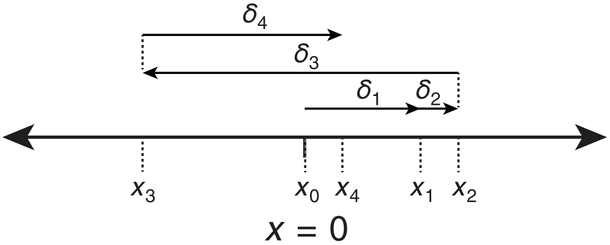
The Gaussian Probability Distribution Function



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

- Mean $= \mu$
- Variance $= \sigma^2$
- Standard deviation $= \sigma$

A 1-dimensional Random Walk with a Distribution of Step Lengths, δ_i



- Assume: n steps and $\langle \delta_i \rangle = 0$.
- $\langle \delta_i^2 \rangle$ unspecified.

Clicker Question #2

For a random walk of n steps (repeated many times) and a variable step size, δ_i, and $\langle \delta_i \rangle = 0$, which of the following are true?

1. $\langle x_n \rangle = 0$
2. $\mathrm{RMS}(x_n) = 0$
3. $\mathrm{RMS}(x_n) = n \langle \delta_i \rangle$
4. $\mathrm{RMS}(x_n) = \sqrt{n} \langle \delta_i \rangle$
5. $\langle x_n^2 \rangle = n \langle \delta_i^2 \rangle$
6. None of the above

A 1-dimensional Random Walk with a Distribution of Step Lengths, δ_i, and $\langle \delta_i \rangle = 0$

- Properties of the distribution of x_n
  - The mean:

$$\mu = \langle x_n \rangle = 0$$

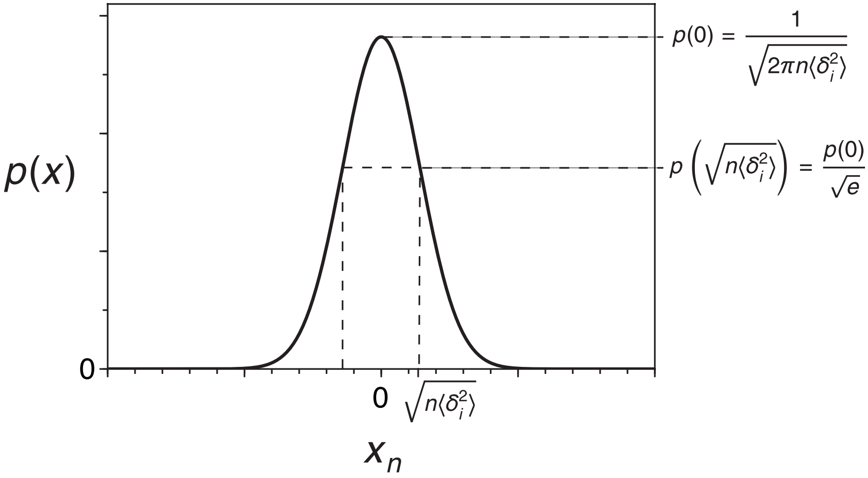
  - The variance:

$$\sigma^2 = \int p(x_n)(x_n - \mu)^2 dx = \int p(x_n)(x_n - 0)^2 dx$$

$$= \int p(x_n) x_n^2 dx = \langle x_n^2 \rangle = n \langle \delta_i^2 \rangle$$

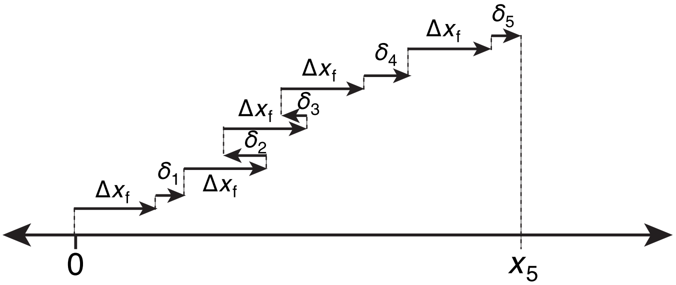
- μ doesn't always equal zero, and σ^2 doesn't always equal $\langle x_n^2 \rangle$ for a random walk.

The Gaussian Distribution of End Points for a 1-dimensional Random Walk with Variable Step Lengths



- Also applies to fixed step lengths, as long as $\langle\delta_i\rangle = 0$.

A Different Kind of Random Walk

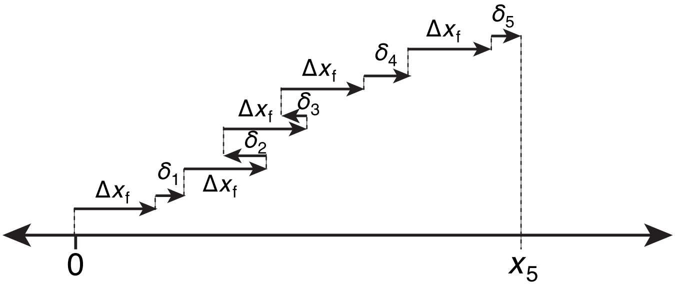


- Each step, i, has two parts:
  1. A fixed displacement of Δx_f, always of the same magnitude and sign.
  2. A random displacement of δ_i, such that:

$$\langle \delta_i \rangle = 0$$

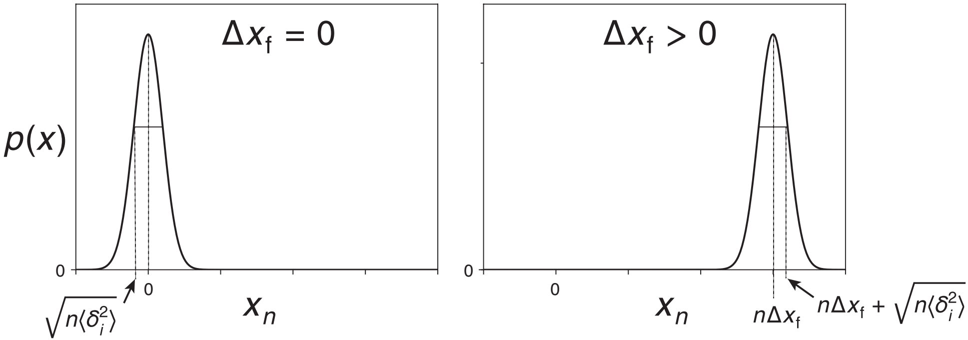
- Each random walk consists of n two-step steps.

A New Kind of Random Walk



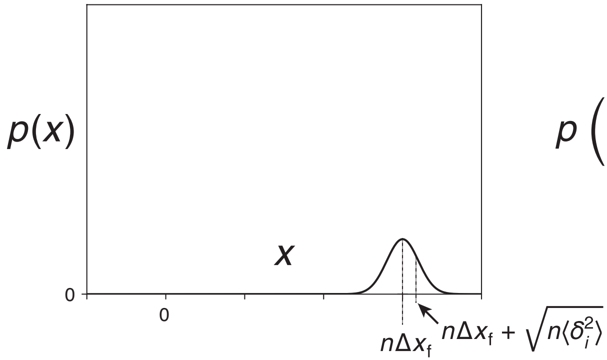
- If $\Delta x_f = 0$, and we do a large number, N, of walks:
  - It's just like our old random walks.
  - $\langle x_n \rangle = 0$
  - $\langle x_n^2 \rangle = n \langle \delta_i^2 \rangle$
  - The values of x_n have a Gaussian distribution, with $\mu = 0$ and $\sigma = \sqrt{n \langle \delta_i^2 \rangle}$.

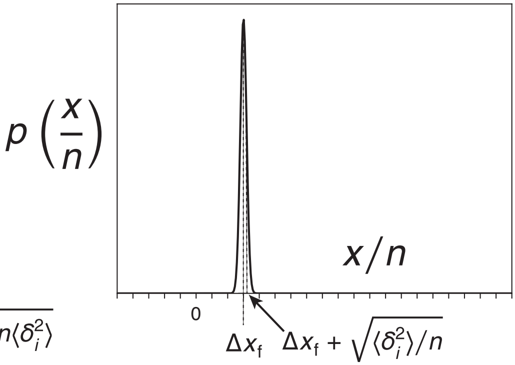
Distribution of Endpoints, x_n, for N Walks



- $\mu = \langle x_n \rangle = n\Delta x_f$
- $\sigma^2 = \langle (x_n - \mu)^2 \rangle = \langle (x_n - n\Delta x_f)^2 \rangle = n\langle \delta_i^2 \rangle$
- A Gaussian distribution, with $\mu = n\Delta x_f$ and $\sigma = \sqrt{n\langle \delta_i^2 \rangle}$.
- As n increases, the distribution moves to the left *and* spreads out.

Suppose that we divide x by n





- $\mu = \langle x_n / n \rangle = \Delta x_f$
- $\sigma^2 = \langle (x_n/n - \mu)^2 \rangle = \langle \delta_i^2 \rangle / n$
- A Gaussian distribution, with $\mu = \Delta x_f$ and $\sigma = \sqrt{\langle \delta_i^2 \rangle / n}$.
- The longer the walk, the narrower the distribution of x/n.

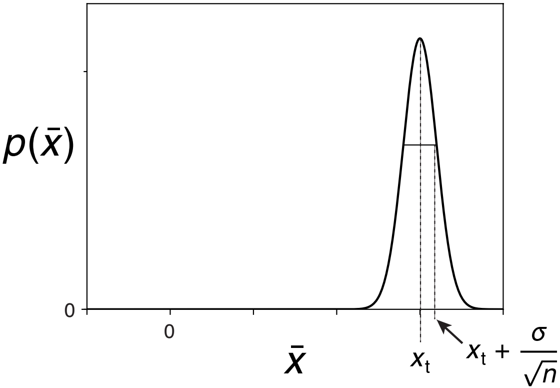
Something Similar to Our Funny Random Walk

- Make n repeated measurements of some physical quantity, x, such as the mass of an object.
- Each measurement, x_i, is the sum of:
  - The "true" value of the measurement, x_t.
  - A random error, δ_i, with $\langle\delta\rangle = 0$ and a standard deviation, $\sigma = \mathrm{RMS}(\delta_i)$.
- Sum all of the measured values:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} (x_\mathrm{t} + \delta_i)$$

- Divide the sum by the number of measurements, n. Call this $\bar{x}$.
- How would we interpret the result: $\bar{x} = \left(\sum x_i\right) / n$?
- What would we expect to find if we were to repeat the series of n measurements a large number of times?

Distribution of Averages of Experimental Measurements



- The most probable value of $\bar{x}$ is x_t, the "true" value of x.
- The width of the distribution is proportional to σ, the standard deviation of the experimental errors.
- The width of the distribution decreases with $\sqrt{n}$.