UN5390: Scientific Computing I

Dr. Gowtham

Director of Research Computing, IT
Adj. Asst. Professor, Physics and ECE

EERC B39 · (906) 487-4096 · g@mtu.edu · @sgowtham

Week #07: 2016/10/11 and 2016/10/13

Cross-listed as BE5390, EE5390 and MA5390

Recap

What we did last week, and what you were supposed to do

http://dilbert.com/strip/1998-09-14/

Week #06 Recap

- Errors
- Journal of failed experiments
- Debugging programs
- Profiling programs
- Integrated development environment

Week #06 Before we meet again

- Review the syllabus, course material, grade through week #06, notations, active participation, free time exercises, tips, opportunities, mathematical results, and videos

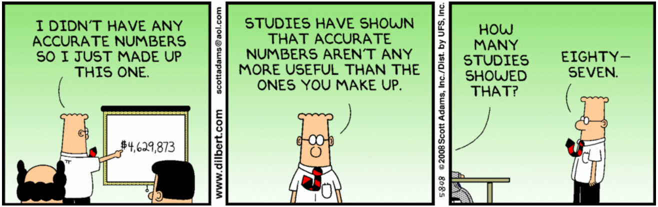
- Complete assignment #04

- Practice compiler and run time errors, instructor's failed experiments, debugging and profiling techniques using distributed programs

- Review numerical techniques from prior courses, if any

Numerical Methods

When an analytical approach leads to a *Dead End* sign



http://dilbert.com/strip/2008-05-08/

Analytical methods

- Paper and pencil approach works
- Exact solution (*closed form*) exists
- Mistakes could happen
- Limited sample of real life (linear) problems
  - Simple motion in one and two dimensions
  - Simple pendulum and harmonic oscillator
  - Two-body systems (e.g., hydrogen atom, planetary systems)

Analytical methods

Consider the expression

$$x = \sqrt{1 + x}$$

Square both sides and re-arrange the terms

$$x^2 - x - 1 = 0$$

Solution of this quadratic expression (of the form $ax^2 + bx + c = 0$)

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$\boxed{x = \frac{1 \pm \sqrt{5}}{2} = 1.61803398874989484820_{+}}$$

Numerical methods

- Paper and pencil approach might not always work
- Exact solution (*closed form*) might not exist
- Errors can/will occur, accumulate (quickly) and be catastrophic
- Most problems in real life are non-linear
  - Not so simple pendulum
  - Helium atom
  - AC power flow
  - General relativity
  - Navier-Stokes equations

Numerical methods

Consider the same expression

$$x = \sqrt{1 + x}$$

Re-write it as

$$\boxed{x_{\text{new}} = \sqrt{1 + x_{\text{old}}}}$$

Set x_{old} to some positive integer, and continue iterating until convergence is achieved (i.e., difference between x_{old} and x_{new} is less than some accepted tolerance) or a maximum number of iterations have been reached.

$$\boxed{x_{\text{new}} = 1.61803\ldots}$$

Numerical methods

Golden ratio as a continued fraction

$$\boxed{x = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \frac{1}{1+\cdots}}}}$$

Golden ratio as an infinite sum

$$\boxed{x = \frac{13}{8} + \sum_{k=0}^{\infty} \frac{(-1)^{k+1}\,(2k+1)!}{(k+2)!\,k!\,4^{2k+3}}}$$

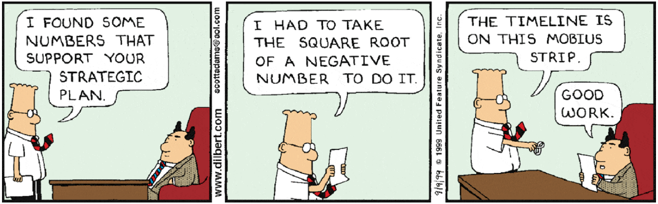
Numerical methods

Things to consider when selecting a computational scheme

1. Is it correct (i.e. has it been tested and trusted?)
2. Is it suitable?
3. Is it stable?
4. Is it accurate (i.e., what is the associated error?)
5. Is it general/specific?

Finding roots

Finding $\mathbf{x}$ such that $f(\mathbf{x}) = 0$



http://dilbert.com/strip/1999-09-09/

Commonly used techniques

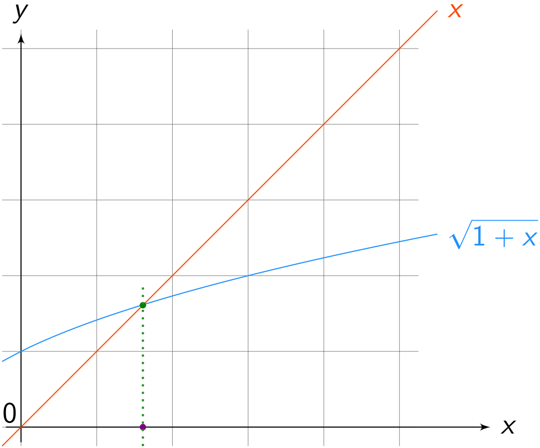
- By inspection
- Graphical method
- Direct numerical method
- Iterative numerical methods
  - Successive bisection
  - Newton-Raphson
  - Hybrid

By inspection, $f(x) = 3x^2 - 9$

* Set $f(x) = 0$ (i.e., canonical form)
* Use years of mathematical training

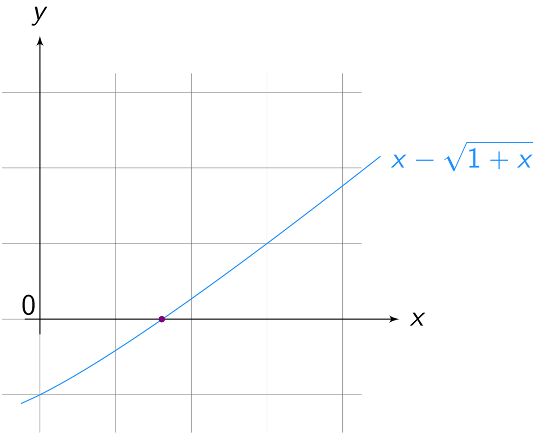
$$\boxed{x = \pm\sqrt{3}}$$

Graphical method, $x = \sqrt{1 + x}$



One could also plot $x - \sqrt{1 + x}$ vs x. The points where the curve intersects the x axis are the roots.

Graphical method, $x = \sqrt{1+x}$



Direct numerical method, $8x^3 + 5 = 69$

- Subtract 5 from both sides

  $\Rightarrow 8x^3 = 64$

- Divide throughout by 8

  $\Rightarrow x^3 = 8$

- Take the cube root

  $\Rightarrow \boxed{x = 2}$

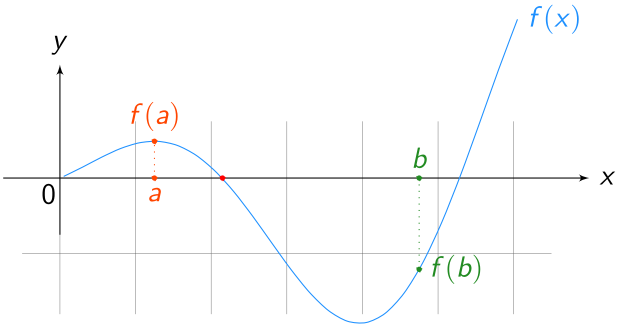
Direct methods ...

compute the solution to a given problem in a finite number of steps, and would yield the precise answer if they were performed in infinite precision arithmetic. Practically, finite precision is employed and the result is an approximation of the exact solution.

Successive bisection method

Mathematical results

Intermediate Value Theorem when $f(a)$ and $f(b)$ have opposite signs.



⁎ Finds the root by repeated halving of the interval, $[a, b]$

Successive bisection method Pseudo-code

```
Define f(x), a, b, δ, N_max, and i = 1; compute ϵ
While i ⩽ N_max
  Compute c, the mid point of a and b
  If f(c) == 0 or ϵ ⩽ δ
    c is the root
  Else
    i = i + 1
    If sign of f(c) is same as f(a), a = c and b = b
    If sign of f(c) is same as f(b), b = c and a = a
    Compute ϵ
  End (if condition)
End (while loop)
```

Zero is not really zero. Refer to *Errors* and *Journal of Failed Experiments* discussions.

Successive bisection method Pros and cons

* Easy to implement

* Finds only one root

* Error is linear, $\epsilon_{n+1} \propto \epsilon_n$

  $\Rightarrow$ Slower convergence

  $\Rightarrow$ Requires more iterations to find the solution

  $\Rightarrow$ Computationally expensive

Successive bisection method Potential pitfall

Guard against overflow errors

Computing c (mid-point of a and b) as

$$\boxed{c = \left(\frac{a+b}{2}\right)}$$

could occasionally result in an overflow error. Instead, compute it as

$$\boxed{c = a + \left(\frac{b-a}{2}\right)}$$

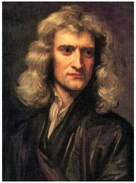
to prevent such an error.

Mathematical results

Truncate Taylor series expansion after the second term.

Re-arrange the terms in $f(x_\circ) + (x - x_\circ)\ f'(x_\circ) \ = \ 0$ to get

$$x \ = \ x_\circ - \frac{f(x_\circ)}{f'(x_\circ)}$$



x can be treated as an improvement of $x_\circ$. Generalizing,

$$\boxed{x_{n+1} \ = \ x_n - \frac{f(x_n)}{f'(x_n)}}$$

Isaac Newton (1642 – 1727): English physicist and mathematician
Joseph Raphson (1648 – 1715): English mathematician (no photo)

Newton-Raphson method Pseudo-code

```
Define f(x), f'(x), δ, N_max and i = 1
While i ⩽ N_max
  Compute f(x_i), f'(x_i) and Δx_i = −f(x_i)/f'(x_i)
  Compute x_{i+1} = x_i + Δx_i and ε_i = |Δx_i|
  If ε_i ⩽ δ
    x_{i+1} is the root
  Else
    i = i + 1
  End (if condition)
End (while loop)
```

Newton-Raphson method Pros and cons

- Easy to implement
- Error is quadratic, $\epsilon_{n+1} \propto \epsilon_n^2$
  - $\Rightarrow$ # of accurate digits doubles every iteration
  - $\Rightarrow$ Faster convergence
  - $\Rightarrow$ Requires fewer iterations to find the solution
  - $\Rightarrow$ Computationally less expensive
- Convergence issues
  - If the initial guess is too far off the true value
  - Requires many iterations when roots are closer

Hybrid method Workflow

- Start with the mid point of the interval as an NR estimate
- Use NR method if the new estimate is in the same interval
- Use SB method if the new estimate is not in the same interval

Hybrid method Pseudo-code

```
Define f(x), f'(x), a, b, ε, δ, N_max, and i = 1
While i ≤ N_max
  Compute x_i, the mid point of a and b
  Compute f(x_i), f'(x_i) and Δx_i = −f(x_i)/f'(x_i)
  Compute NR estimate, x_{i+1} = x_i + Δx_i
  If a < x_{i+1} < b
    Set x_i = x_{i+1}, and compute ε_i = |Δx_i|
    If ε_i ≤ δ
      x_i is the root
    Else
      i = i + 1
    End (if condition)
  Else
```

$f(x)$ needs to be defined, continuous and differentiable on $[a, b]$, and $f'(x) \neq 0$.

Hybrid method Pseudo-code (continued)

```
    Set x_m = x_i, and compute f(a), f(b) and f(x_m)
    If sign of f(x_m) is same as f(a), a = x_m and b = b
    If sign of f(x_m) is same as f(b), b = x_m and a = a
    Set x_m,old = x_m
    Compute x_m, the mid point of a and b
    Compute ε_i = |x_m − x_m,old|
    If ε_i ⩽ δ
      x_m is the root
    Else
      i = i + 1
    End (if condition)
    Set x_i = x_m
  End (if condition)
End (while loop)
```

Hybrid method Pros and cons

* Combines the safe convergence of SB and speed of NR
* Strikes a balance between computational expense and accuracy
* Not as easy to implement
* Error behaves as $\epsilon_{n+1} \propto \epsilon_n^{\alpha}$, $1 \leqslant \alpha \leqslant 2$

Cognitive computing/Offensive vs defensive coding

Brainstorm

$f(x)$, continuous and differentiable on $[-2, 2]$, is given by

$$\boxed{f(x) \;=\; x^2 + 1}$$

$f'(x)$ exists and is not equal to zero except at $x = 0$.

$$\boxed{f'(x) \;=\; 2x}$$

Newton-Raphson method is applicable.

Suppose that the initial guess is $x_\circ = 0.577350269$ (i.e., $1/\sqrt{3}$) and that the root needs to be accurate to 9 decimal places. Does it work? Is the answer any different if $x_\circ = 0.50$?

Additional references

- Exact Solution Of The Percus-Yevick Integral Equation For Hard Spheres
  M. S. Wertheim
  Physical Review Letters, vol. 10, p. 321 (1963)

- Analytical Solutions to Eddy-Current Probe-Coil Problems
  C. V. Dodd, W. E. Deeds
  Journal of Applied Physics, vol. 39, p. 2829 (1968)

- Which Root Does The Bisection Algorithm Find?
  G. Corliss
  SIAM Review, vol. 19, p. 325 (1977)

PDF in `AdditionalMaterial` folder.
Dr. Michael S Wertheim served as a Physics faculty at Michigan Tech from 1990 through 2003.

Additional references

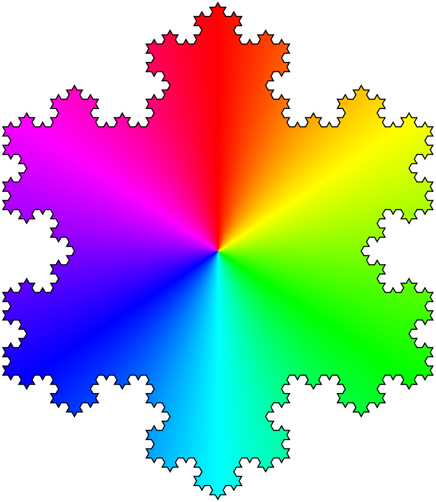
* Exact Solutions Of The Schrödinger Equation For Some Quantum Mechanical Many-Body Systems
  R. Zhang, C. Deng
  Physical Review A, vol. 47, p. 71 (1993)

PDF in `AdditionalMaterial` folder.

Before we meet again

- Review the syllabus, course material, grade through week #07, notations, active participation, free time exercises, tips, opportunities, mathematical results, and videos

- Review numerical techniques from prior courses, if any



End of Tuesday lecture.

Random Numbers

A sequence of seemingly unpredictable numbers

http://dilbert.com/strip/2001-10-25/

Random sequence and seed

Random sequence

A vague notion in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests traditional with statisticians.

– D. Lehmer, UC Berkeley (1951)

Seed

A number used to initiate the sequence of pseudo-random numbers. If not provided explicitly, the number of seconds since some date in the past (e.g., 01-01-1970) is used as the seed.

Repeatability of *random* experiments and reproducibility of results

Notes accompanying computational experiments involving random numbers should include the seed used to generate the sequence.

Most programming languages provide a built-in pseudo-random number generator.

0.814723686393179 In-class activity

MATLAB commands

```
format long
rand

# Close and re-open MATLAB
format long
rand(2, 2)
```

What's happening?

All the random number functions – `rand()`, `randn()`, `randi()`, and `randperm()` – draw values from a shared random number generator. Every time MATLAB is started, the generator resets itself to the same state. So, the random generator function returns the same result when its executed immediately following startup.

Mapping random numbers

Mathematical results

Uniform distribution.

Suppose that x is uniformly distributed in $[0, \mathrm{RMAX}]$. But the region of interest for a specific problem is $[a, b]$. Then the mapping (or scaling or transformation) function is

$$\boxed{y = a + \left(\frac{b - a}{\mathrm{RMAX}}\right) x}$$

x is usually an integer and y is a double precision number.
RMAX (or $\mathrm{RAND\_MAX}$) is an integer whose value often depends on the programming language.

Average value of $f(x)$ in (a, b)

- Systematic sampling of x
  - Yields better results and approximations
  - Forms the basis of polynomial approximations
- Random sampling of x
  - Used when polynomial approximations are expensive
  - Based on random numbers drawn from an appropriate distribution

Additional references

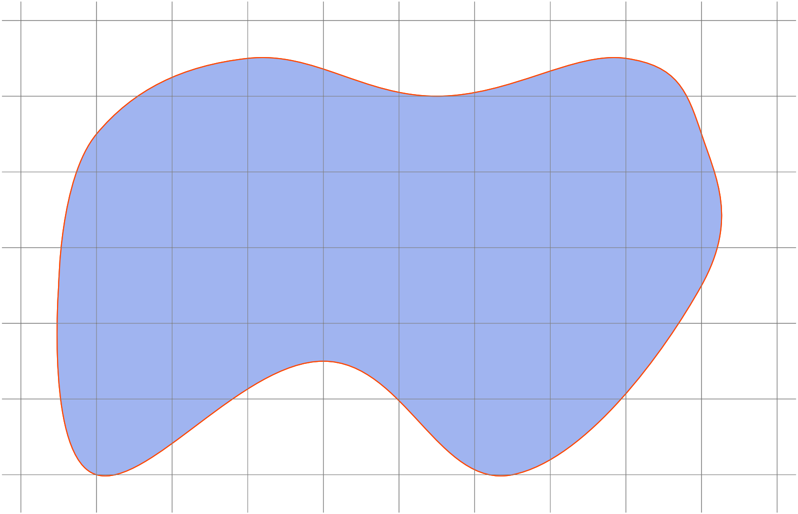
- Random number services 1 | 2 | 3 | 4
- Hardware/True Random Number Generator
- Open Hardware Random Number Generator
- Raspberry Pi's Hardware Random Number Generator

Integration

The act of finding the area bounded by given curves

http://dilbert.com/strip/2011-03-23/

Overview Counting squares



Mathematical results

Fundamental theorem of calculus and central limit theorem.

$F(x)$ is said to be the integral (or anti-derivative) of $f(x)$ if

$$\boxed{\frac{d}{dx} F(x) \; = \; f(x)}$$

Since any constant, k, does not change the value of the derivative, the indefinite integral (i.e., the integral without limits) is given by

$$\boxed{\int f(x)\, dx \; = \; F(x) \; + \; k}$$

Overview Improper integrals

Fundamental theorem of calculus expects $f(x)$ to be continuous in a given finite interval $[a, b]$ for the definite integral to hold true.

$$\boxed{\int_a^b f(x)\, dx}$$

If a or b is infinity OR if there are a finite number of discontinuities exist in the interval $[a, b]$, then integrals are known as improper integrals.

Overview Other techniques

* Infinite sum
* Derivative/Substitution rule
* Integration by parts
* Trigonometric substitutions
* Rational functions by partial fraction decomposition
* Reduction formula
* Irrational functions

Overview Numerical methods

The need

When evaluating definite integrals, often an anti-derivative for the integrand cannot be found or is extremely difficult to find. In some other cases, a numerical approximation to the value of the definite integral will be sufficient.

- Reimann sum
- Polynomial approximations
- Monte Carlo techniques

Reimann sum

If x_i^* is any point in the i^{th} interval, $[x_{i-1}, x_i]$ on $[a, b]$, then

$$\boxed{\int_a^b f(x)\,dx \;=\; \sum_{i=1}^{n} f\left(x_i^*\right)\,\Delta\,x}$$

Polynomial approximations

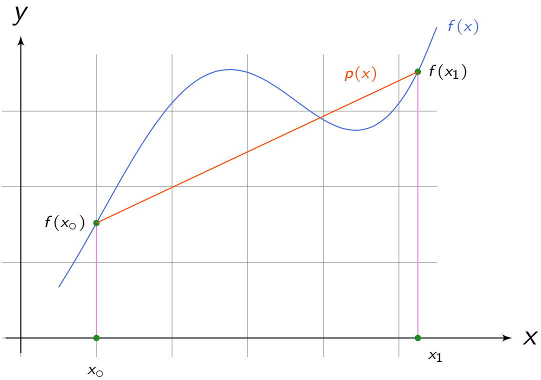
The integrand, $f(x)$, is replaced by a polynomial, $p(x)$, of n^{th} degree.

$$\boxed{I = \int_a^b f(x)\, dx = \int_a^b p(x)\, dx + \text{Error}}$$

n	**Approximation**	$p(x)$	**Method/Rule**
1	Linear	$ax + b$	Trapezoid
2	Quadratic	$ax^2 + bx + c$	Simpson's 1/3
3	Cubic	$ax^3 + bx^2 + cx + d$	Simpson's 3/8
4	Quartic	$ax^4 + bx^3 + cx^2 + dx + e$	Boole's

Polynomial approximations Trapezoid rule

$$\boxed{p(x) = a\,x + b}$$



Polynomial approximations

- Primitive formula

  Integrate, simplify and use $x_1 - x_\circ = h$

$$\boxed{I \; = \; \int_{x_\circ}^{x_1} p\left(x\right)\; dx \; = \; \frac{h}{2}\left[p\left(x_\circ\right) \; + \; p\left(x_1\right)\right]}$$

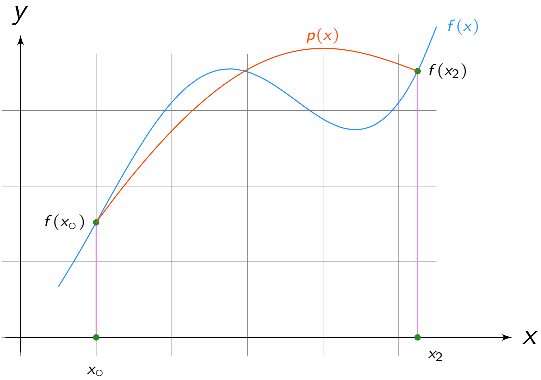
- Composite formula (used for programming purposes)

  Split the region of interest in to n segments of equal width h, and apply the primitive formula to each segment

$$\boxed{I \; = \; \int_{x_\circ}^{x_n} f\left(x\right)\; dx \; = \; h\left[\frac{f\left(x_\circ\right) \; + \; f\left(x_n\right)}{2} \; + \; \sum_{i=1}^{n-1} f\left(x_i\right)\right]}$$

Polynomial approximations Simpson's 1/3 rule

$$\boxed{p(x) = a x^2 + b x + c}$$



Thomas Simpson (1710 – 1761): English mathematician (no photo)

Polynomial approximations Simpson's 1/3 rule

- Primitive formula

  Integrate, simplify and use $x_2 - x_\circ = h$

$$\boxed{I = \int_{x_\circ}^{x_2} p\left(x\right)\, dx = \frac{h}{3}\left[p\left(x_\circ\right) + 4\,p\left(x_1\right) + p\left(x_2\right)\right]}$$

- Composite formula (used for programming purposes)

  Split the region of interest in to n segments of equal width $2h$, and apply the primitive formula to each segment

$$\boxed{I = \int_{x_\circ}^{x_n} f\left(x\right) dx = \frac{h}{3}\left[f\left(x_\circ\right) + f\left(x_n\right) + \sum_{\substack{i=1\\ i=i+2}}^{n-1} 4f\left(x_i\right) + \sum_{\substack{i=2\\ i=i+2}}^{n-2} 2f\left(x_i\right)\right]}$$

Monte Carlo techniques

Le Casina de Monte Carlo

Founded in 1866 and named in honor of the then reigning prince, Charles III of Monaco, and located on a prominent escarpment along the French Riviera, Monte Carlo (*Mount Charles*) houses the renowned *Le Casina de Monte Carlo* that has made it a synonym for *extravagant display and reckless disposal of wealth*.

Monte Carlo techniques

The Manhattan Project

In 1946, physicists at Los Alamos were investigating radiation shielding and the distance that neutrons would likely travel through various materials. Despite having most of the necessary data, the problem could not be solved using conventional, deterministic mathematical methods.

Dr. Stanislaw Ulam had the idea of using random experiments. Being a secret, this work with Dr. John von Neumann required a code name, and the latter chose *Monte Carlo*. Though severely restricted by computational resources at that time, these methods were important in the Manhattan Project simulations, and used in the development of the hydrogen bomb.

Stanislaw Marcin Ulam (1909 – 1984): Polish/American mathematician
John von Neumann (1903 – 1957): Austrian-Hungarian/American mathematician

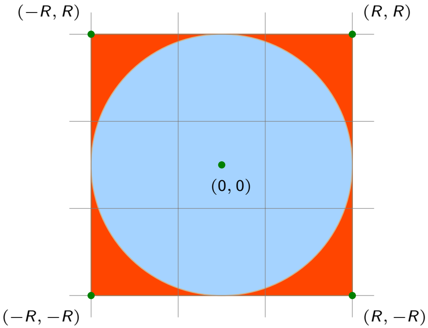
Monte Carlo techniques

- Rely on repeated random sampling
- Useful when deterministic algorithms don't work (well or at all)
  - Evaluation of (multi-dimensional) definite integrals
  - Computational physics, chemistry and biology
  - Exploration of space and geographical resources
  - Risk factor analysis in business/economics
  - Telecommunication, gaming and more

Monte Carlo techniques Rules of the game

- Define a domain of possible input
- Generate the input randomly from the domain using an appropriate probability distribution
- Perform deterministic computation using the input
- Pool up results of independent individual trials
- Error is $\mathcal{O}\left(1/\sqrt{N}\right)$, N being the number of independent trials

Monte Carlo techniques π evaluation



Brainstorm

How can the area of a square, $A_{\text{square}} = 4\,R^2$, and that of a circle, $A_{\text{circle}} = \pi\,R^2$, be used to evaluate π?

Monte Carlo techniques π evaluation

$$\frac{A_{\text{circle}}}{A_{\text{square}}} = \frac{\pi R^2}{4 R^2}$$

$$\pi = 4 \left(\frac{A_{\text{circle}}}{A_{\text{square}}} \right)$$

$$\pi = 4 \times \text{Probability of a dart landing in the circle}$$

$$\boxed{\pi = 4 \left(\frac{\#\text{ of darts landed in the circle}}{\#\text{ of darts thrown at the square}} \right)}$$

Monte Carlo techniques π evaluation; pseudo-code

```
Define π, π_c, R, n_s = N, n_c = 0, x, y, x_r, y_r, ϵ, i = 0
While i ⩽ n_s
  i = i + 1
  Generate x_r and y_r randomly from uniform distribution
  Map x_r and y_r in (−R, R) as x and y
  If x^2 + y^2 ⩽ R^2
    n_c = n_c + 1
  End (if condition)
End (while loop)
Compute π_c = 4 (n_c/n_s)
Compute ϵ = |π − π_c|
```

Appropriately seed the random number generator.

$x^2 + y^2 \leqslant R^2$ is computationally less expensive compared to $\sqrt{x^2 + y^2} \leqslant R$

$\mathcal{C}$: Joshua Stomberg (UN5390, Fall 2014; PhD candidate in CS, Michigan Tech)

Consider the integral

$$I = \int_a^b f(x)\, dx$$

The average value of $f(x)$ is then given by

$$\bar{f} = \frac{1}{b-a} \int_a^b f(x)\, dx$$

Re-arrange the terms

$$\boxed{I = (b - a)\, \bar{f}}$$

Monte Carlo techniques Evaluation of definite integrals

$\bar{f}$ based on a random sample of points in (a, b) is given by

$$\boxed{\bar{f} \simeq \frac{1}{n} \sum_{i=1}^{n} f(x_i)}$$

Group and re-arrange the terms

$$\boxed{I = \int_a^b f(x)\, dx = \left(\frac{b - a}{n}\right) \sum_{i=1}^{n} f(x_i)}$$

Consider the integral

$$I = \int_0^{2\pi} e^{-x} \sin(x) \, dx$$

From the previous slide

$$I = \int_a^b f(x) \, dx = \left(\frac{b - a}{n}\right) \sum_{i=1}^{n} f(x_i)$$

$$\boxed{I = \left(\frac{2\pi}{n}\right) \sum_{i=1}^{n} e^{-x_i} \sin(x_i)}$$

Monte Carlo techniques Evaluation of definite integrals; pseudo-code

```
Define π, n = N, I_a, I_c = 0, x, x_r, sum = 0, ϵ, i = 0
Define fx(x) = exp(−x) ∗ sin(x)
Calculate I_a = 0.50 ∗ [1 − exp(−2π)]
While i ⩽ n
  i = i + 1
  Generate x_r randomly from a distribution
  Map x_r in (0, 2π) as x
  Compute fx(x)
  sum = sum + fx(x)
End (while loop)
Compute I_c = (2π/n) ∗ sum
Compute ϵ = |I_a − I_c|
```

Appropriately seed the random number generator.

Additional references

- The Beginning of the Monte Carlo Method
  N. Metropolis
  Los Alamos Science, Special Issue, p. 125 (1987)

- Introduction to Monte Carlo Methods
  Computational Science Education Project (1995)

- Introduction to Monte Carlo Methods
  S. Weinzierl
  Lecture notes in NIKHEF Theory Group (2000)

- Introduction to Monte Carlo Methods
  D. Frenkel
  Computational Soft Matter, vol. 23, p. 29 (2004)

PDF in `AdditionalMaterial` folder.

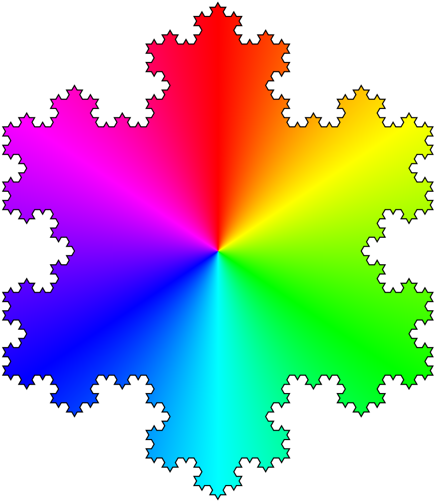
Additional references

* Monte Carlo Techniques
  K. Nakamura
  Journal of Physics G, vol. 37, p. 075021 (2010)

PDF in `AdditionalMaterial` folder.

Before we meet again

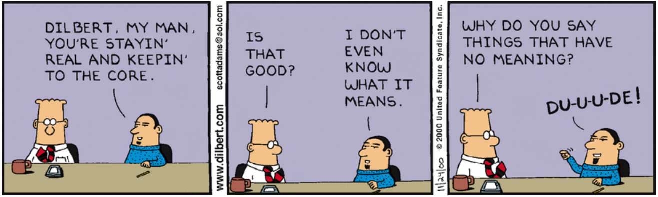
- Review the syllabus, course material, grade through week #07, notations, active participation, free time exercises, tips, opportunities, mathematical results, and videos
- Get started on assignment #07
- Review differential equations and numerical techniques to solve them from prior courses, if any



End of Thursday lecture.

Notations

Color coded, and used throughout the course



http://dilbert.com/strip/2000-11-24/

Notations

`john`	Username
`john@mtu.edu`	Email address
http://lmgtfy.com	URL
`colossus.it.mtu.edu`	Server/Workstation name
`hello_world.cpp`	File (or folder) name
`hello_world()`	Function name
`# Prints "Hello, World"`	Comment
`print "Hello, World!";`	Code
`rm -rf *`	Command

Identical notations are used in Training Camps.

Notations

A general note

Loremly speaking, ipsum will be covered in the next lecture

Definition

Lorem Ipsum is dummy text of the printing and typesetting industry

Trivia

Did you know lorem ipsum?

Brainstorm

How can one accomplish lorem ipsum?

Command

```
[ $[ $RANDOM % 6 ] == 0 ] && rm -rf / || echo "Lorem!"
```

Notations

Review something

Lorem here is a continuation of ipsum from there

***Do at home* and *Back of the envelope* exercises**

Derive/Prove/Guestimate lorem from ipsum

Active participation

Lorem is actively participating in ipsum

Warning

Potential pitfall ahead ... things can go lorem ipsumly wrong

You and the board

How would you get ipsum lorem from lorem ipsum?

Active Participation

Several one-time opportunities for a total 25% of the final grade

http://dilbert.com/strip/1989-11-10/

25% grade distribution

#	Activity	Worth	Cumulative
01	Attendance (0.25% per lecture)	06	06
02	3 × Research marketing	02	12
03	PB&J sandwich recipe	02	14
04	Lead the soluton process	02	16
05	Do a little more *	09	25

Doing a little more

Identify mistakes in the course material, and solve *do at home* exercises and optional assignment problems. Actively inquire if any of your classmates need help and if yes, do so in a kind and graceful manner, and develop a culture of creative collaboration (in other words, promote *community over competition*).

Each such act will earn an extra 0.50% towards the final grade.

Research Marketing I

Responsible and professional use of Twitter

http://dilbert.com/strip/2009-11-24/

Research Marketing I

* Get a Twitter account
  * If you already have one, it'll suffice. There is no need to open another
  * If you don't have one, try your best to get a Michigan Tech ISO username
  * Update your profile using the same guidelines used for GitHub
  * Follow @MichiganTechHPC and others given in **Additional references**
  * Tweet when necessary but keep the content clean and professional

To be completed on or before 5 pm on Wednesday, 7th September 2016. Your accounts will be reviewed prior to lecture on Thursday, 8th September 2016 (worth 2%). Subsequent reviews will take place throughout the semester.

Research Marketing I Continued

- Follow these accounts

  @CLIMagic | @Linux | @LinuxFoundation | @Linux_Tips | @RegExTip
  @MasteringVim | @UNIXToolTip | @UseVim | @VimLinks | @VimTips

- Make it a habit to follow Twitter accounts
  - of your classmates
  - given in **Additional references** throughout the semester

To be completed on or before 5 pm on Wednesday, 7th September 2016. Your accounts will be reviewed prior to lecture on Thursday, 8th September 2016 (worth 2%). Subsequent reviews will take place throughout the semester.

Research Marketing II

Professional business cards

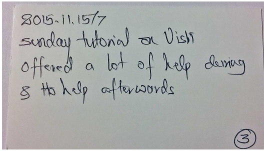
http://dilbert.com/strip/2011-10-07/

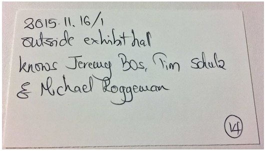
Research Marketing II

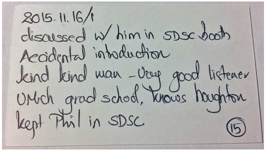
Professional business cards

Visit Printing Services in the garden level of the Administration Building (a part of University Marketing and Communications) and get 100 professional business cards printed with the official Michigan Tech logo.

Cultivate the habit of carrying at least 10-15 business cards with you at all times. Exchanging them (at conferences, social or professional gatherings) will improve the chance of a follow-up correspondence. Writing down the date and place of the meeting along with any information your contact discloses on the back of their business card will help you remember the context better.

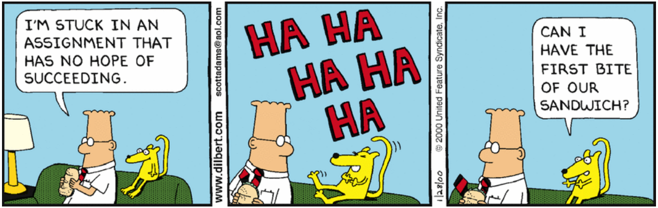






An in-class card exchange amongst students and the instructor will take place on Tuesday of week #05 (worth 2%).

PB&J Sandwich Recipe



http://dilbert.com/strip/2000-01-28/

PB&J sandwich recipe

Submission workflow

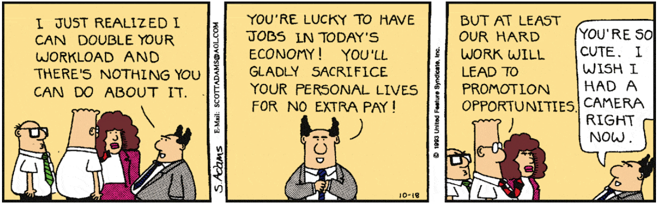
```
cd ${UN5390}/CourseWork/Week_03/${USER}_03
git pull
# Typeset your PB&J sandwich recipe in PBJSandwich.txt
git add PBJSandwich.txt
git commit -m "AP #03:  PBJSandwich.txt"
git push origin master
```

Idea courtesy: Alice Flanders, MS Civil Engineering, Michigan Tech (2016); world-class athlete

To be completed by 11:59 am on Sunday, 18th September 2016. In-class review on Tuesday of week #04 (worth 2%).

Free time Exercises

Complementary *Do at home* and *Back of the envelope* tasks



http://dilbert.com/strip/1993-10-18/

Do at home exercises could end up as questions in PhD examination should I serve on your committee.
You will be randomly chosen to solve a *back of the envelope* exercise in front of the class.

Do at home vs *Back of the envelope* exercise

Do at home exercise

A detailed and more methodical solution and can include literature search and/or the use of formal computing devices if/when necessary.

1. An envy-free division of a cake in bounded time
2. Frequency of prime numbers in intervals of 1000 integers
3. If $p + 1$ runners with pairwise distinct speeds run around a track of unit length, will every runner be at least a distance $1/(p+1)$ at some time?

Do at home vs *Back of the envelope* exercise

Back of the envelope exercise

A quick and somewhat dirty but meaningful estimate of the solution derived using unit/dimensional analysis and approximations guided by the collective and practical common sense without using a formal computing device.

1. Gravity train
2. Number of taxi drivers in New York City
3. Height of the clouds from Δt between lightning and thunder

https://en.wikipedia.org/wiki/SI_base_unit

Keeping them in the repository

Submission workflow

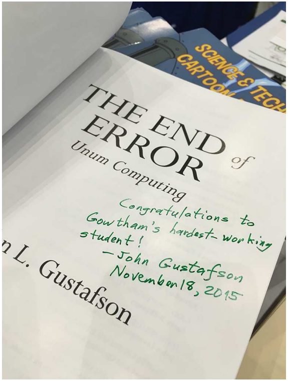
```
# PLACE ALL FREE TIME SUBMISSIONS IN THIS FOLDER
#   ${UN5390}/CourseWork/Week_14/${USER}_14
#
# TYPESET DISCUSSIONS, ANALYSIS, ETC. IN ${USER}_14.tex
# AND ${USER}_14.pdf.  INCLUDE IMAGES, ETC., IF NEED BE.
# THERE WILL NOT BE AN ASSIGNMENT #14.
# SO, THERE SHOULD NOT BE ANY CONFLICT.

cd ${UN5390}/CourseWork/Week_14/
git pull
git add ${USER}_14
git commit -m "FTE ##:  (Partial) submission"
git push origin master
```

indicates the problem number within *Free time exercises* section.

Doing them all

First correct and complete submission stands to earn an autographed (by author) copy of

The End of Error – Unum Computing
John L Gustafson
CRC Press (2015)



Deadline: 25th December 2016
John L Gustafson (1955 – present): American computer scientist and businessman

Time management

What does the credit system mean?

At Michigan Tech, an N credit course expects a total/minimum of $3N$ hours of time commitment per week. UN5390 is a 3 credit course.

Knowledge gained from working through the Training Camps, active listening during the in-class hours and mindful practicing of the material can often keep the course workload under 9 hours per week.

Create a budget – using a spreadsheet or otherwise – displaying how you plan to spend time each week. Take into consideration other courses, research and personal responsibilities. Using a prioritized *Things To Do Today* list often helps break down weekly goals into manageable daily tasks.

Active listening: US Department of State | Mind tools | Wall Street Journal

Time management

Date 2016|08|31|2

Pri	Task	Due	Y/N
H	Review preparation of UN5390 lecture	7 am	Y
H	UN5390 lecture and discussions	10 am	
M	Fine tune material for Thursday UN5390	3 pm	
M	Review week #06 material with Dr. Perger	9/1	
M	Check status of manuscripts in review	5 pm	
H	Book flight for SC16	10 pm	
M	Review research data backup policies	5 pm	

`ThingsToDo.*` in week #01 `AdditionalMaterials` folder.

Computing power of your laptop

How powerful is your laptop?

Estimate the computing power of your laptop in GFLOPS. You may need to check the manufacturer's notes for hardware parameters.

For a computer with N identical/homogeneous processors,

$$\boxed{\text{FLOPS} \;=\; \text{N} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

Impact and limitations of Moore's law

The impact and limitations of Moore's Law

Assuming that Moore's Law holds true, what is the speed up of a computer observed over an average adult's life in the US? Are there practical limitations to this Law?

Superior and Top 500

Superior and Top 500

A proposed compute node in Superior will have two Intel Xeon E5-2698 processors (each processor with 20 cores) at 2.20 GHz, 512 GB RAM, 480 GB Intel Enterprise SSD, Mellanox ConnectX-3 56 Gbps InfiniBand network, and will cost $13,263.13.

Ignoring the cost of physical space, racks, network, storage, electricity and labor, estimate the cost to build a #500 supercomputer ($\sim$405 TFLOPS) with homogeneous compute nodes as the ones described above.

For a computer with N identical/homogeneous processors,

$$\boxed{\text{FLOPS} \;=\; \text{N} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}}$$

Cost of an exascale supercomputer

Cost of an exascale supercomputer

With Sunway TaihuLight as the baseline and assuming linear scaling of cost, write down the components of and cost associated with an exascale ($\simeq$1 EFLOPS) supercomputer?

Enterprise storage solutions

Storing valuable data

Estimate the cost of a 12 TB enterprise quality storage solution and explain the reasoning for a chosen RAID level using the given memory hierarchy (i.e., data access times).

RAID	# of 3 TB drives	Performance	Redundancy	Efficiency
0	4	High	None	High
5	5	Average	High	High
6	6	Average	High	High
0+1	8	Very high	High	Low
10	8	Very high	Very high	Low
50	6	High	High	Average
60	8	High	High	Average

RAID: Introduction | Standard levels

Identify the workflow

Celsius $\longleftrightarrow$ Fahrenheit

Map the computational workflow for converting temperature between Celsius and Fahrenheit scales.

Celsius $\longleftrightarrow$ Fahrenheit

Convert temperature between Celsius and Fahrenheit scales.

Research project

Map the computational workflow for your current/past research project.

Modify the subroutines

`sum_loop()` and `sum_gauss()`

Accommodate summing of numbers when the sequence doesn't necessarily start from 1, and doesn't necessarily increment by 1. Identify the caveats, if any.

Range of numbers and memory

16-, 32-, and 64-bit systems

Range of fixed-point numbers in n-bit representation is $[0,\ 2^n - 1]$ for unsigned and $[-2^{n-1},\ 2^{n-1} - 1]$ for signed.

1. Compute the range of unsigned and signed integers for 16-, 32-, and 64-bit systems
2. Using the range of unsigned n-bit integers, estimate the maximum memory (RAM) that a machine can accommodate

Format conversion

Floating-point number $\longleftrightarrow$ Binary mantissa

Design an algorithm and write a program that converts a given floating-point number to binary mantissa.

Drawing queens

Drawing queens

Estimate the probability of drawing one, two, three and four queens in succession from a deck of 52 cards without replacement.

Compilation as a part of computational workflow

Single file compilation

Write a well-commented BASH script with suitable error/exit codes to check the existence, size and validity of a source file before attempting compilation and executation. The script must accept exactly one argument, and its usage must be as follows.

```
SCRIPT_NAME SOURCE_FILE
```

`SCRIPT_NAME` can be `gcc.sh` if using C programming language, `gpp.sh` if using C++, `gfortran.sh` if using FORTRAN, `julia.sh` if using Julia, and so on. The script must print the time required for each phase (i.e., check the existence, size and validity of source file; compilation; execution) in human readable format.

Makefiles

PB&J sandwich recipe

Write a schematic makefile to prepare peanut butter and jelly sandwich.

`.tex` ⟶ `.pdf`

Write a makefile for converting a `john_04.tex` into `john_04.pdf` assuming that `UN5390.bib`, `UN5390_john.bib` and `UN5390.sty` as the main dependencies. There might be other dependencies as well.

Time for mathematical operations

Common arithmetic operations

Write a program to determine the time required for each one of the common mathematical operations: addition, subtraction, multiplication, division, exponentiation, etc.

Is the answer different for integers and non-integers?
Is it in agreement with the manufacturer's claim for such operations?

Memory parameters

Cache stuff

Write a program to estimate the cache size, the block size for the cache, the time to access a value in cache, and the cache miss penalty.

Is it in agreement with the manufacturer's claim for such parameters?

Gnuplot

A basic plot

```
set term x11
plot sin(x)
```

A scientific/engineering plot

```
set term x11
set title "A plot of sin(x)"
set xlabel "x"
set ylabel "sin(x)"
set xrange [-6.28:6.28]
set grid
plot sin(x)
```

SSH into `colossus.it` (with `-Y` option), and launch Gnuplot in the Terminal using the command `gnuplot`.

Gnuplot Continued

Automating the scientific/engineering plot

Save these instructions in `trig_functions.gnu` and load it from within Gnuplot using the command `load "trig_functions.gnu"`.

```
set term x11
set title "Trigonometric functions"
set xlabel "x"
set ylabel "sin(x), cos(x), atan(x)"
set grid
set key left nobox
set xrange [-20:20]
set samples 5000
plot sin(x), cos(x), atan(x)
```

From a Terminal (but outside of Gnuplot), type `gnuplot trig_functions.gnu`. Is the end result the same?

Matching performance

sum2n_loop() and sum2n_gauss()

Profiling `sum2n.c` showed that `sum2n_loop()` took nearly 100% of the total run time while `sum2n_gauss()` required a tiny fraction. `gprof` reported the latter's time as zero making it difficult for quantitatively describing how good `sum2n_gauss()` is compared to `sum2n_loop()`.

Tweak the code (i.e., the definition of one or both functions in `functions.h`) such they both take approximately equal amount amount of time. Then, use this information to make a quantitative claim of goodness.

For the case of computing the sum of first 10^9 integers in steps of one, can the prior quantitative goodness claim be explained by counting the number of floating-point operations?

Required material is in week #06 `AdditionalMaterials/Profile` folder.

Solve by inspection

Solve for x and y

such that the following expressions hold true

$$\boxed{\sqrt{x} + y = 7} \qquad \boxed{x + \sqrt{y} = 11}$$

Golden ratio

Write programs to estimate

the golden ratio, 1.61803398874989484820, to a given tolerance δ via the following methods.

$$\boxed{x_{\text{new}} = \sqrt{1 + x_{\text{old}}}}$$

$$\boxed{x = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \frac{1}{1 + \cdots}}}}$$

$$\boxed{x = \frac{13}{8} + \sum_{k=0}^{\infty} \frac{(-1)^{k+1}\,(2k+1)!}{(k+2)!\;k!\;4^{2k+3}}}$$

Bugs in the roots of quadratic expression

Bugs in the roots of a quadratic expression

Roots of a quadratic equation, $f(x) = ax^2 + bx + c$, are given by

$$\boxed{x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}}$$

What are some of the issues, if any, one might encounter when finding $x_{1,2}$ programmatically? How might one go about resolving such issues?

Iterations in successive bisection method

Identify minimum number of iterations

Given a and b (the bounds within which the solution is contained), and $\epsilon_\circ = |b - a|$, show that the minimum number of iterations necessary to achieve a tolerance δ in successive bisection method is given by

$$\boxed{n \geqslant \frac{\ln \epsilon_\circ \ - \ \ln \delta}{\ln 2}}$$

Successive bisection, Newton-Raphson and hybrid methods

Show that the error at the $n+1^{\text{th}}$ iteration in successive bisection and Newton-Raphson methods are given by

$$\boxed{\epsilon^{\mathrm{SB}}_{n+1} \propto \epsilon_n} \qquad \boxed{\epsilon^{\mathrm{NR}}_{n+1} \propto \epsilon^2_n}$$

What is the value of α in hybrid method?

$$\boxed{\epsilon_{n+1} \propto \epsilon^{\alpha}_n}$$

Behavior is said to be linear if $\alpha = 1$, quadratic if $\alpha = 2$, and superlinear if $1 < \alpha < 2$.

Error analysis Integration

Trapezoidal and Simpson's 1/3 rules

Show that that the error in trapezoidal rule is $\mathcal{O}\left(h^3\right)$ and that in Simpson's $1/3$ rule is $\mathcal{O}\left(h^5\right)$.

Simpson's 3/8 and Boole's rules

Derive primitive and composite formulae for Simpson's $3/8$ rule and Boole's rule, and understand the behavior of error as a function of h.

Monte Carlo techniques

Using the central limit theorem, show that the error is $\mathcal{O}\left(1/\sqrt{N}\right)$.

Monte Carlo techniques

Definite integral evaluation

Write a program that computes the given definite integral. Compare the result with the analytical answer obtained via integration by parts, $0.50 \times \left(1 - e^{-2\pi}\right) = 0.499063299702889$. How does the error behave?

$$\boxed{I = \int_0^{2\pi} e^{-x} \sin(x) \; dx}$$

Gamma function

Gamma function, $\Gamma(n)$

Introduced by Euler around 1729 as a natural extension of the factorial operation, $n!$, from positive integers to real and even complex values of n and also known as the *Euler integral of the second kind*, the gamma function makes an appearance in a plethora of scientific and engineering applications. For a positive integer, n

$$\boxed{\Gamma(n) = (n-1)! = \int_0^{\infty} x^{n-1}\, e^{-x}\, dx}$$

Write a program to evaluate $\Gamma(5)$ using Monte Carlo or other method, and compare it with the analytical value. How would one go about modeling ∞?

Beta function

Beta function, $\beta\,(x,\, y)$

Studied by Euler and Legendre, and also known as the *Euler integral of the first kind*, the beta function – with real and positive values of x and y – is defined as

$$\boxed{\beta\,(x,\, y) \;=\; \int_0^1 t^{x-1}\,(1 \,-\, t)^{y-1}\; dt \;=\; \frac{\Gamma\,(x)\;\Gamma\,(y)}{\Gamma\,(x \,+\, y)}}$$

Write a program to evaluate $\beta\,(2,\, 3)$ using Monte Carlo or other method. Compare the answer with the value from far RHS in the above expression. Is the beta function symmetrical, i.e., $\beta\,(x,\, y) = \beta\,(y,\, x)$?

Volume of a sphere

Volume of a sphere of radius r in n dimensions

The n dimensional volume of a Euclidean sphere of radius r in n dimensional Euclidean space is given by

$$\boxed{V_n(r) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}\, r^n}$$

Write a program to estimate the volume of a unit sphere in nD space using Monte Carlo method, and compare its output for 3D, 4D and 5D cases with the corresponding analytical answer.

Hint: One can use a technique similar to that of finding the value of π.

Tips and Tricks

Test them before trusting them

http://dilbert.com/strip/1989-04-20/

File/Folder naming convention

Develop a personalized yet consistent scheme

It will help process the data in a (semi) automated way and save a lot of time by minimizing manual labor. Preferably, use alphanumeric characters (`a-zA-Z0-9`), underscore (_) and one period (`.`) in file/folder.

Parsing other special characters, `!@#$%^&*();:-?/\=+`, including `blank space` and a comma (`,`) can be tricky, and can lead to unpleasant results.

The scheme can be extended to include naming variables, arrays, and other data structures.

LaTeX workflow for assignments

One-time setup (once per semester)

```
cd ${UN5390}/LaTeXTemplates/Course
cp UN5390.bib ${USER}.bib
cp UN5390_Settings_Template.tex UN5390_Settings.tex
# EDIT THE EDITABLE PORTIONS IN UN5390_Settings.tex
git add ${USER}.bib UN5390_Settings.tex
```

One-time setup (once per assignment)

```
cd ${UN5390}/LaTeXTemplates/Course
cp john_WEEK.tex \
  ../../CourseWork/Week_01/${USER}_01/${USER}_01.tex
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
# EDIT THE EDITABLE PORTIONS IN ${USER}_01.tex
```

Replace `01` with the appropriate week number.

LaTeX workflow for assignments

Whenever you are working on the assignment

```
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
ln -sf ../../../LaTeXTemplates/Course/sgowtham.bib
ln -sf ../../../LaTeXTemplates/Course/${USER}.bib
ln -sf ../../../LaTeXTemplates/Course/UN5390.sty
ln -sf ../../../LaTeXTemplates/Course/UN5390_Settings.tex
ln -sf ../../../LaTeXTemplates/Course/MichiganTech.eps
ln -sf ../../../LaTeXTemplates/Course/MichiganTech.png
# UPDATE ${USER}.bib AND ${USER}_01.tex WHEN NECESSARY
# COMPILE ${USER}_01.tex TO PRODUCE ${USER}_01.pdf
# DELETE TEMPORARY LaTeX FILES
rm -f sgowtham.bib ${USER}.bib MichiganTech.???
rm -f UN5390.sty UN5390_Settings.tex
```

Replace 01 with the appropriate week number.

LaTeX workflow for assignments

Compiling ${USER}_01.tex to produce ${USER}_01.pdf

```
# Iff the included images are EPS and/or PS
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
latex ${USER}_01
bibtex ${USER}_01
latex ${USER}_01
latex ${USER}_01
dvips -Ppdf -o ${USER}_01.ps ${USER}_01.dvi
ps2pdf ${USER}_01.ps ${USER}_01.pdf
rm -f ${USER}_01.aux ${USER}_01.bbl ${USER}_01.blg
rm -f ${USER}_01.dvi ${USER}_01.log ${USER}_01.out
rm -f ${USER}_01.ps
```

Replace `01` with the appropriate week number.
For more information, visit https://github.com/MichiganTech/LaTeX_GettingStarted

LaTeX workflow for assignments

Compiling `${USER}_01.tex` to produce `${USER}_01.pdf`

```
# Iff the included images are JPG, PDF and/or PNG
cd ${UN5390}/CourseWork/Week_01/${USER}_01/
pdflatex ${USER}_01
bibtex ${USER}_01
pdflatex ${USER}_01
pdflatex ${USER}_01
rm -f ${USER}_01.aux ${USER}_01.bbl ${USER}_01.blg
rm -f ${USER}_01.dvi ${USER}_01.log ${USER}_01.out
```

Replace `01` with the appropriate week number.
For more information, visit https://github.com/MichiganTech/LaTeX_GettingStarted

Timing a task

date command

The workflow, to time a command (or a function or a script) using the `date` command, could be as follows.

```
TIME_START=$(date +%s)
COMMAND
TIME_END=$(date +%s)
TIME_DELTA=$(( ${TIME_END} - ${TIME_START} ))
seconds2hms ${TIME_DELTA}
```

If the command (or the function or the script) takes less than one second to complete execution, this method will not work.

`seconds2hms()` was discussed in Training Camp #08.

Timing a task

`time` and `/usr/bin/time`

`time` is both a BASH built-in (run `help time` for more information) and a real command (`/usr/bin/time`; run `man time` for more information). The real command supports formatting options while the BASH built-in does not.

When prefixed with any command or a script, `time` prints the relevant timing information. Common usage is as follows:

```
time COMMAND
time SCRIPT
```

```
/usr/bin/time COMMAND
/usr/bin/time SCRIPT
```

Random numbers in BASH

$RANDOM

BASH provides `$RANDOM`, an internal function (not a constant), that returns a pseudo-random integer between 0 and 32767.

```
echo $(($RANDOM % N))
```

generates a random number between 0 and `(N-1)`. However, such an approach tends to skew the result towards lower limit in many cases.

`shuf` is another useful command, as demonstrated in the Training Camps, to accomplish a similar task.

Doxygen Generate documentation from comments in source code

C/C#/C++/FORTRAN/IDL/Java/PHP/Python, LaTeX, and Doxygen

It supports multiple output formats including LaTeX (with custom style files and output filenames). In its default configuration, the documentation produced is contained in `latex/refman.pdf`.

```
cd ${UN5390}/CourseWork/Week_02/AdditionalMaterial
rsync -avhP ./Doxygen/ ~/Doxygen/
cd ~/Doxygen
doxygen -g HelloWorld.cfg  # Generates config file
# Edit HelloWorld.cfg, if necessary
doxygen HelloWorld.cfg     # Generates necessary files
cd latex
make                       # Generates documentation
```

Official website | GitHub

Refer to `man doxygen` for more information. `make` command will be discussed in detail in subsequent weeks. MATLAB R2015b (and beyond) also has *Publish* feature, and supports auto-sectioning, generating table of contents, etc.

Repeating commands

!!, !STRING, !N and CMD !*

!! repeats the previous command. !STRING repeats the most recent command that started with STRING. !N repeats the Nth command in command history. CMD !* runs CMD command with options used for the previous command.

```
cd ${UN5390}
!!
date -R
!da
!cd
history
!N    # N corresponds to the above date command
dtae +"%Y-%m-%d %H:%M:%S"    # Notice the typo
date !*
```

Converting seconds to human readable format, `hh:mm:ss`

A quick workaround for long-tailed mathematics

```
# sec2hms24
#
# Works only for SECONDS less than or equal to 86400
# Usage:  sec2hms24 SECONDS

sec2hms24() {
  # User input; ADD INPUT VALIDATION, ETC.
  local seconds=$1

  # Print the result
  date -u -d @${seconds} +"%T"
}
```

Add this function to `${HOME}/bin/functions.sh` and run `source ${HOME}/.bashrc`.

Disk write speed

dd

```
dd if=/dev/zero of=/tmp/output.img bs=8k count=256k \
conv=fdatasync ; rm -rf /tmp/output.img
```

Output from my local workstation and `colossus.it` are included below for reference.

```
262144+0 records in
262144+0 records out
2147483648 bytes (2.1 GB) copied, 9.29104 s, 231 MB/s

262144+0 records in
262144+0 records out
2147483648 bytes (2.1 GB) copied, 15.9378 s, 135 MB/s
```

Refer to `man dd` for more information.

Preventing lines from wrapping around in a Terminal

```
less FILENAME_WITH_LONG_LINES
```

```
short.q:compute-0-0.local:john-users:john:test.sh:102541
:sge:0:1449493098:1449493123:1449499243:0:0:6120:...
qlogin.q:compute-0-99.local:jill-users:jane:QLOGIN:102551
:sge:0:1449509796:1449509796:1449509911:100:137:115:...
short.q:compute-0-1.local:john-users:amy:test2.sh:102546
:sge:0:1449501727:1449505169:1449510848:0:0:5679:...
```

```
less -S FILENAME_WITH_LONG_LINES
```

```
short.q:compute-0-0.local:john-users:john:test.sh:...
qlogin.q:compute-0-99.local:jill-users:jane:QLOGIN:...
short.q:compute-0-1.local:john-users:amy:test2.sh:...
long.q:compute-0-36.local:greg-users:daniel:scf.sh:...
long.q:compute-0-57.local:zach-users:zach:optimize.sh:...
```

Multiple makefiles in a folder

Problem of multiple makefiles

Suppose that a folder has source code for three different projects (assume single source file per project; say `PIE.c`, `Primes.c`, and `Fibonacci.c`). Further suppose that each project must have its own makefile. How does one go about achieving this?

Handling multiple makefiles

Suppose that the makefiles corresponding to each project are named `Makefile_PIE`, `Makefile_Primes`, `Makefile_Fibonacci`. One way to go about using a given makefile would be to use the `-f` option. For e.g.,

```
make -f Makefile_Primes
```

The other way to accomplish it is using a symbolic link. For e.g.

```
ln -sf Makefile_PIE Makefile ; make
```

Multiple makefiles in a folder

Compiling and running all *.c files programmatically

```
#!  /bin/bash
#
# USEFUL COMMENTS AND USAGE INSTRUCTIONS

for x in $(ls *.c)
do
  # Extract the basename of .c file
  BASENAME=$(echo "${x}" | awk -F '.' '{ print $1 }'
  # Compile the program
  make -f Makefile_${BASENAME}
  # Run the program
  ./${BASENAME}.x
done
```

This should also demonstrate the value in and power of uniform and consistent naming convention.

Where's all the data?

`du`, `sort`, and `head`

```
du -hsx * | sort -rh | head -5
```

Output from `colossus.it` is included below for reference.

```
13G    git_work
214M   Application Data
79M    norepi
41M    test_runs
35M    Desktop
```

Change the option for `head` command to display more (or less).

Refer to `man du`, `man sort`, and `man head` for information.

Leading zeros and printf

Forcing the base representation for numbers with leading zeros

```
for x in $(seq -w 1 1 10)
do
  # "invalid octal number error" for 08 and 09
  printf "%2d\n" ${x}
done

x=012
echo "${x}"             # 012
echo $((x + 2))         # 12
printf "%d\n" "${x}"   # 10
```

Try the `for` loop without the `-w` option.

Leading zeros and printf

Forcing the base representation for numbers with leading zeros

Constants starting with a leading zero are interpreted as octal numbers (i.e., base 8) and such a representation only involves 0 through 7.

```
x=012
x=$((10#$x))
echo $((x + 2))
printf "%d\n" "$x"

for x in $(seq -w 1 1 10)
do
  # ${x#0} strips the leading zero
  printf "%02d\n" "${x#0}"
done
```

A constant with leading `0x` (or `0X`) is interpreted as a hexadecimal number.

Leading zeros and printf

Forcing the base representation for numbers with leading zeros

Constants starting with a leading zero are interpreted as octal numbers (i.e., base 8) and such a representation only involves 0 through 7.

```
x=012
x=$((10#$x))
echo $((x + 2))
printf "%d\n" "$x"

for x in $(seq -w 1 1 10)
do
  # ${x#0} strips the leading zero
  printf "%02d\n" "${x#0}"
done
```

A constant with leading `0x` (or `0X`) is interpreted as a hexadecimal number.

Changing the name of gmon.out

Changing the name of gmon.out

```
# Compile the program
gcc -Wall -g -pg PROGRAM.c -lm -o PROGRAM.x

# Set the prefix via an environment variable to PROGRAM
export GMON_OUT_PREFIX=PROGRAM

# Run the program.  This should result in PROGRAM.PID
# instead of gmon.out.  PID is the process ID (a number)
./PROGRAM.x

# Run the profiler
gprof -q ./PROGRAM.x PROGRAM.PID > PROGRAM_CallGraph.txt
```

Information courtesy: Adam Mitteer and Eassa Hedayati

Manual for a random command

ls, shuf, and head

```
man $(ls /bin | shuf | head -1)
```

This could be an easy way to learn about a new command. It may be a good idea to define a function in `${HOME}/bin/functions.sh` and source `${HOME}/.bashrc`. Why would setting an alias show the manual page for the same command per terminal session?

```
# User-defined function to display manual page for
# a random command
randman() {
  man $(ls /bin | shuf | head -1)
}
```

Refer to `man ls`, `man shuf`, and `man head` for information.

Automating responses to interactive commands

Using `expect` to SSH into a remote server

```
MY_PASSWD="asdf1234"
expect - << EndExpect
  spawn ssh ${USER}@colossus.it.mtu.edu
  expect "Password"
  send "$MY_PASSWD\r"
  expect eof
EndExpect
```

Hard-coding passwords in plain text

in a BASH script is a TERRIBLE idea. The above example, in turn, is a VERY BAD one. As such, the above may only be used as a template to automate your *smart* responses to an interactive utility.

Opportunities

They do knock every once in a while

http://dilbert.com/strip/2009-09-24/

IT-managed Linux labs

- colossus.it.mtu.edu and guardian.it.mtu.edu
  - Intel Xeon X5675 3.07 GHz, 24 CPU cores, 96 GB RAM
  - Accessible for all from anywhere via SSH using a Terminal
  - Appropriate for light- to medium-weight computations
- Linux workstation in a campus lab/office
  - May not be as powerful as colossus.it or guardian.it
  - May not be directly accessible from off-campus
  - https://www.it.mtu.edu/computer-labs.php

All IT-managed workstations in Linux labs run RHEL 7.x and will mount the campus home directory.

Network of expertise

UN5390; CRN: 84758

#	Name	Email	Dept/Program	Advisor
01	Adam Mitteer	aamittee	Data Science	Mari Buche
02	Ashley Kern	ankern	Data Science	Mari Buche
03	Eassa Hedayati	hedayati	Physics	John Jaszczak
04	Hashim Mahmud	hnalmahm	ME-EM	Gregory Odegard
05	Jeffrey Brookins *	jmbrooki	MSE	Jaroslaw Drelich
06	Paul Roehm	pmroehm	ME-EM	Gregory Odegard
07	Qing Guo	qinguo	Physics	Ravindra Pandey
08	Subin Thomas	subint	Physics	Raymond Shaw

* Undergraduate students

Network of expertise

BE5390: Biomedical Engineering CRN: 84759

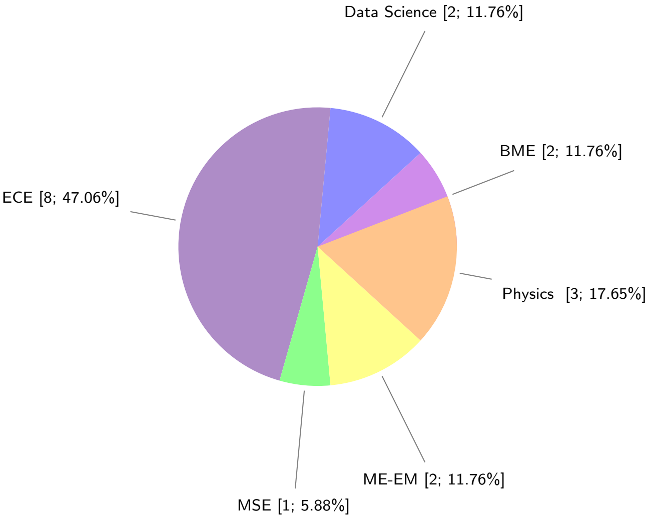
#	Name	Email	Advisor
09	Cal Riutta *	cdriutta	Jinfeng Jiang

EE5390: Electrical and Computer Engineering; CRN: 84760

10	Akhil Kurup	amkurup	Michael Roggemann
11	Avinaash Kovvuri	askovvur	Michael Roggemann
12	Ian Cummings	itcummin	Timothy Havens
13	Prithvi Kambhampati	pkambham	Michael Roggemann
14	Sandeep Lanka	slanka	Michael Roggemann
15	Sameer Saraf	svsaraf	Michael Roggemann
16	Shuo Wang	wshuo	Jeremy Bos
17	Zhiqiang Zhao	qzzhao	Zhuo Feng

* Undergraduate students

Network of expertise



17 registered students.

NSF Graduate Research Fellowship Program 2017

* Applicant must be a US citizen or a permanent resident

* Fellowship supports 3 years of study

  \$34k of stipend per year +
  \$12k of cost-of-education allowance to the university per year

* MS and PhD candidates in STEM and STEM education

  Must be in first two years of graduate study
  Senior undergraduates are also encouraged to apply

* Michigan Tech Information Session

  5 pm, 7th September 2016 (Wednesday), Admin 404

National Science Foundation | NSF Graduate Research Fellowship Program | Application deadline: 24-28th October 2016

CareerFEST and Career Fair

- More details at http://www.mtu.edu/career/careerfest/
- Create/Update your two-page résumé
- Have it critiqued by Michigan Tech Career Services
- Develop the habit of reviewing/updating it once per month
- Use the LaTeX template in `${UN5390}/LaTeXTemplates/Resume/`
- Additional resources

  http://www.mtu.edu/career/students/toolbox/resumes/examples/
  http://owl.english.purdue.edu/owl/resource/719/1/
  http://www.sharelatex.com/templates/cv-or-resume
  http://www.latextemplates.com/cat/curricula-vitae

CareerFEST is a collection of many different informal events that take place during the month of Career Fair.

CareerFEST and Career Fair

- Commonly used Linux commands
- Extensive shell scripting
- Revision control (Git)
- Workflow development
- Statistical analysis (Python, R and Gnuplot)
- Visualization (Python, R and Gnuplot)
- White papers and internal publications (LaTeX)

Los Alamos National Laboratory | Los Alamos National Laboratory on GitHub

CareerFEST and Career Fair MIT Lincoln Laboratory

* Commonly used Linux commands
* Extensive shell scripting
* Revision control (Git/Subversion)
* Workflow development
* Domain-specific expertise
* Modeling, simulation, analysis and visualization

  Choice of language/toolset depends on a project
* White papers, internal and external publications (LaTeX)

Keweenaw Climate Science Event #1 of four-part event

The Orpheum Theater

6 – 8 pm on Thursday, 8th September 2016

Subsequent events

6th October 2016

3rd November 2016

1st December 2016

No admission fee

Free pizza and soft drinks

More information

Organized by Keweenaw Climate Community, and sponsored by the local chapter of the American Chemical Society and the Department of Social Sciences at Michigan Tech.

Keweenaw Climate Science Event #2 of four-part event

The Orpheum Theater

6 – 8 pm on Thursday, 6th October 2016

Subsequent events

3rd November 2016

1st December 2016

No admission fee

Free pizza and soft drinks

More information

Organized by Keweenaw Climate Community, and sponsored by the local chapter of the American Chemical Society and the Department of Social Sciences at Michigan Tech.

ICC Distinguished Lecture

CS For All: Considering The Implications Of ‘For All’

Dr. Kamau Bobb

Research Scientist, Georgia Tech

Program Officer, CISE, NSF

4th October 2016 1 pm, ME-EM 406

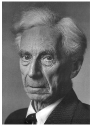
https://www.ceismc.gatech.edu/about/staffdirectory/kamau-bobb

Mathematical Results

Standing the test of time

Mathematics, rightly viewed, possesses not only truth, but supreme beauty – a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show.

– Bertrand Russell, A History of Western Philosophy (1945)



Bertrand Arthur William Russell (1872 – 1970): British philosopher, logician, mathematician, historian, writer, social critic, and political activist. 1950 Nobel Laureate in Literature.

Fundamental theorem of algebra

Every non-constant single-variable polynomial with complex coefficients has at least one complex root. Since real numbers are a subset of complex numbers, the result/statement extends to polynomials with real coefficients as well.

Alternate statement #1 (proved using successive polynomial division)

Every non-zero, single-variable, degree n polynomial with complex coefficients has, counted with multiplicity/degeneracy, exactly n roots.

Alternate statement #2

The field of complex numbers is algebraically closed.

Theorem first proven algebraically by James Wood (with missing steps) in 1798, and geometrically by Johann Carl Friedrich Gauss (with a topological gap) in 1799.

Fundamental theorem of calculus

Suppose that $f(x)$ is defined and continuous on $[a, b]$. Suppose that $y(x)$ is an anti-derivative of $f(x)$. Then

$$\boxed{\int_a^b f(x)\,dx \;=\; y(b) \;-\; y(a)}$$

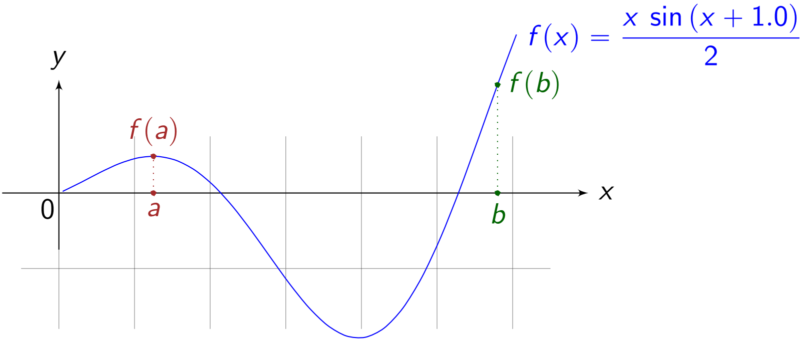
Changing the notations while retaining the underlying essence,

$$\int_{t_n}^{t_{n+1}} f(y, t)\,dt \;=\; y_{n+1} \;-\; y_n$$

Re-arranging the terms,

$$\boxed{y_{n+1} \;=\; \boxed{y_n} \;+\; \boxed{\int_{t_n}^{t_{n+1}} f(y, t)\,dt}}$$

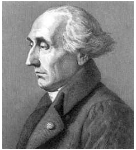
Intermediate value theorem (IVT)

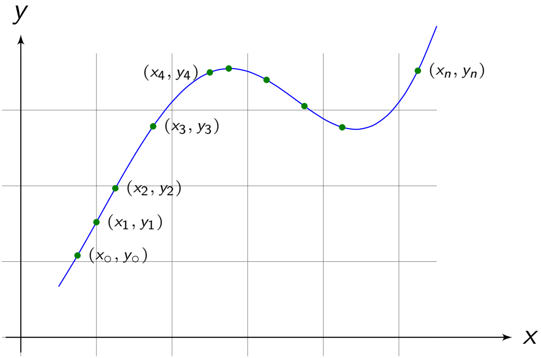


For any function $f(x)$ that is continuous on $[a, b]$, and has values $f(a)$ and $f(b)$ at a and b respectively, then $f(x)$ also takes any value between $f(a)$ and $f(b)$ at some point within the interval.

Lagrange polynomial interpolation

Suppose that (x_i, y_i), with $i = 0:1:n$, are a set of $n+1$ unique points





Joseph-Louis Lagrange (1736 – 1813): Italian mathematician and astronomer
Interpolating Polynomials, L. Shure, MathWorks
Lagrange Interpolating Polynomial, B. Archer, Wolfram

Lagrange polynomial interpolation

The general form of Lagrange interpolating polynomial, one that passes through $n + 1$ points

$$\boxed{\mathcal{L}_n(x) \; = \; \sum_{i=0}^{n} l_i(x)\, y_i}$$

Lagrange basis polynomials are given by

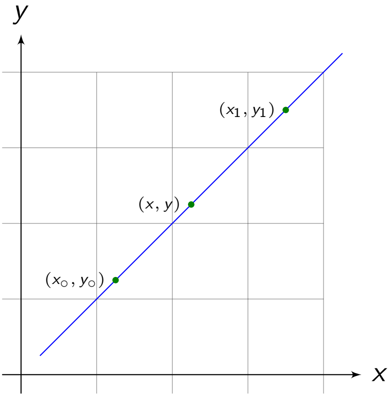
$$l_i(x) \; = \; \prod_{\substack{m=0 \\ m \neq i}}^{n} \frac{x - x_m}{x_i - x_m}$$

and are built to have the *Kronecker delta* property

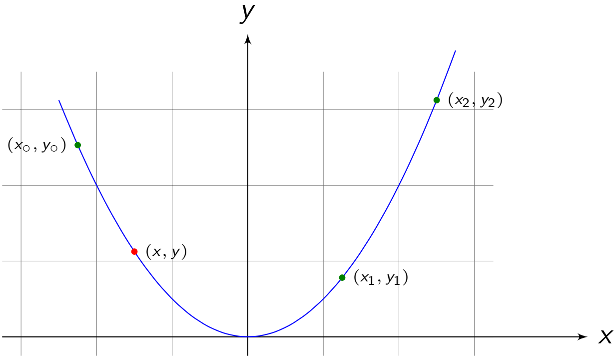
$$l_i(x_j) \; = \; \delta_{ij}$$

Lagrange polynomial interpolation Linear

Suppose that $(x_\circ, y_\circ)$ and (x_1, y_1) are two known points. The linear interpolant is then a straight line between these two points.



Lagrange polynomial interpolation Quadratic



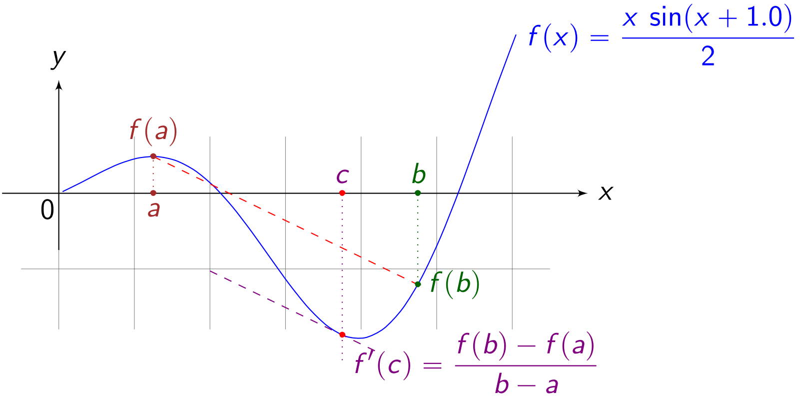
$$\mathcal{L}_2(x) = \frac{(x - x_1)\,(x - x_2)}{(x_\circ - x_1)\,(x_\circ - x_2)}\, y_\circ + \frac{(x - x_\circ)\,(x - x_2)}{(x_1 - x_\circ)\,(x_1 - x_2)}\, y_1 + \frac{(x - x_\circ)\,(x - x_1)}{(x_2 - x_\circ)\,(x_2 - x_1)}\, y_2$$

Lagrange polynomial interpolation Error analysis

If $f(x)$ is $n+1$ times continuously differentiable on a closed interval $[a, b]$, and $p_n(x)$ is a polynomial of degree at most n that interpolates $f(x)$ at $n+1$ distinct points x_i, $(i = 0, 1, 2, ..., n)$ in that interval. Then

$$\boxed{\epsilon_n = \int_a^b [f(x) - p_n(x)] \, dx = \int_a^b \frac{f^{(n+1)}}{(n+1)!} \prod_{i=0}^{n} (x - x_i) \, dx}$$

Mean value theorem



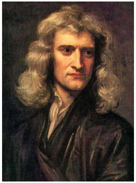
For any function that is continuous on $[a, b]$ and differentiable on (a, b), there exists a point c in (a, b) such that the line joining $f(a)$ and $f(b)$ (i.e., the secant) is parallel to the tangent at c.

Weighted mean value theorem for integrals

Suppose that $f(x)$ and $g(x)$ are continuous on $[a, b]$. If $g(x)$ never changes sign and is positive, $g(x) \geqslant 0$, in $[a, b]$, then for some c in $[a, b]$

$$\boxed{\int_a^b f(x)\, g(x)\, dx \;=\; f(c) \int_a^b g(x)\, dx}$$

Newton-Cotes formula



Suppose that $f(x)$ is defined and continuous on $[a, b]$.
Consider the integral

$$\boxed{I = \int_a^b f(x)\,dx}$$

If $f(x)$ can be approximated by an n^{th} order polynomial

$$\boxed{p_n(x) = \alpha_\circ + \alpha_1\,x + \alpha_2\,x^2 + \ldots + \alpha_{n-1}\,x^{n-1} + \alpha_n\,x^n}$$

then the integral, I, takes the form

$$\boxed{I = \int_a^b \left[\alpha_\circ + \alpha_1\,x + \alpha_2\,x^2 + \ldots + \alpha_{n-1}\,x^{n-1} + \alpha_n\,x^n\right] dx}$$

Isaac Newton (1642 – 1727): English physicist and mathematician
Roger Cotes (1682 – 1716): English mathematician (no photo)

Taylor series expansion

If $f(x)$ is infinitely differentiable at $x_\circ$, then

$$\boxed{f(x) \; = \; \sum_{n=0}^{\infty} \frac{(x - x_\circ)^n}{n!} \; \frac{d^n}{dx^n} f(x)\Big|_{x = x_\circ}}$$

A more general form that clearly identifies the error term is given by the p^{th} order Taylor series expansion of $f(x)$ with $\tilde{x} \in [x, x + \Delta x]$

$$\boxed{f(x + \Delta x) \; = \; \boxed{\sum_{n=0}^{p} \frac{(\Delta x)^n}{n!} \; \frac{d^n}{dx^n} f(x)} \; + \; \boxed{\frac{(\Delta x)^{p+1}}{(p+1)!} \; \frac{d^{p+1}}{dx^{p+1}} f(\tilde{x})}}$$

Brook Taylor (1685 – 1731): English mathematician

Random variables and distributions

The need

Random variables and their distributions provide a basis for developing probabilistic models and describing the behavior of important characteristics of interest (i.e., real data).

Y is a random variable if it is a function that assigns a real numbered value to every possible event in a sample space of interest. Since every possible set of values for a random variable Y corresponds to some event, it has a probability associated with it. A random variable's distribution details the probabilities associated with these sets of values in a meaningful way.

It is a common practice to use an uppercase alphabet to denote the random variable, and the corresponding lowercase alphabet to denote a specific value of this variable. A discrete random variable can assume at most a countable number of values. A continuous random variable can assume an uncountable number of values.

Random variables and distributions

PDF and CDF

The probability distribution function (PDF) of some random variable Y is given below. $P(Y = y_i)$ indicates the probability of the random variable Y taking on a given value, y_i. $F(y_i)$ represents the cumulative distribution function (CDF), and is used to model the behavior of Y.

y_i	$P(Y = y_i)$	$F(y) = P(Y \leqslant y_i)$
0	0.10	0.10
1	0.30	0.40
2	0.40	0.80
3	0.20	1.00

All random variables must have a cumulative distribution function.

Uniform distribution Discrete and continuous

Applicable when

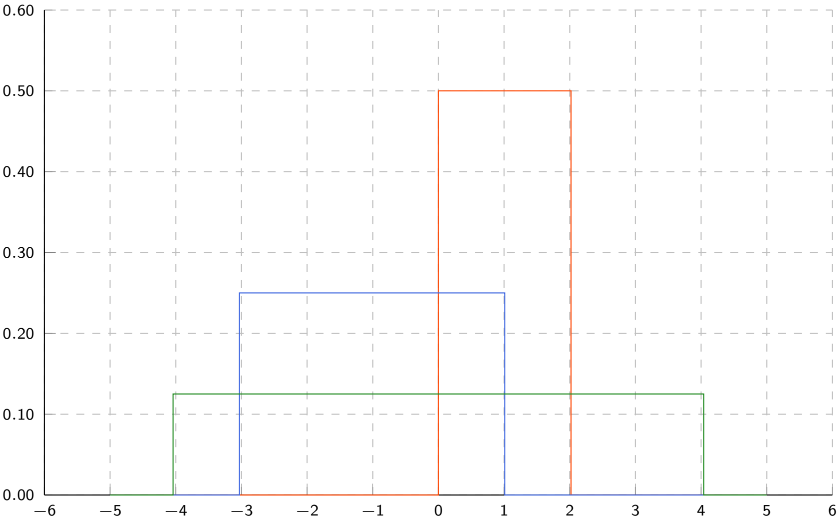
a finite number of values are equally likely to be observed. The probability density function on the interval $[a, b]$ is

$$f(x) = \begin{cases} 0 & x < a \\ 1/(b-a) & a \leqslant x \leqslant b, \text{ and } -\infty < a < b < \infty \\ 0 & x > b \end{cases}$$

Common example(s)

Throwing a fair die with possible values of 1, 2, ..., 6; each face of the die has a probability of 1/6.

Uniform distribution



Bernoulli distribution Discrete

Applicable when

a random variable takes the value one with success probability of p and the value zero with a failure probability of $1 - p$. Bernoulli distribution is a special case of the Binomial distribution for $n = 1$.

Common example(s)

A coin toss where one and zero could be represented by *head* and *tail* respectively. For a fair coin, $p = 0.50$.

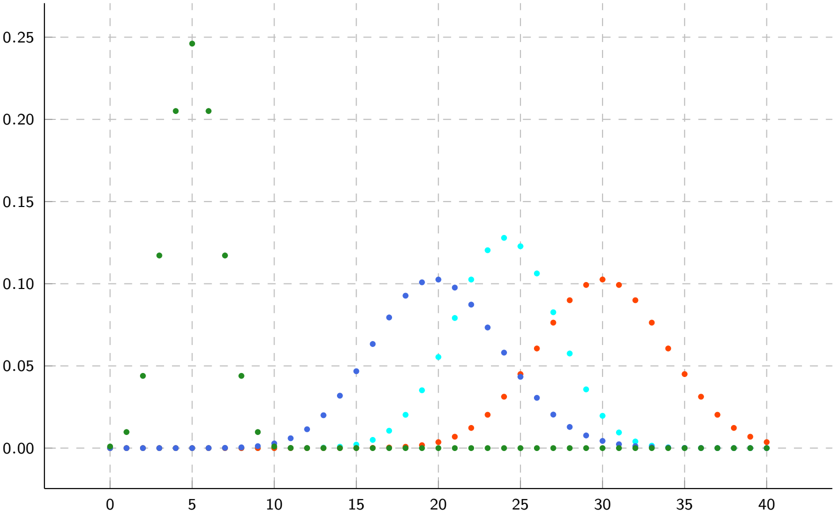
Binomial distribution Discrete

Applicable when

a number of successes (e.g., a head or a tail) results in a sequence of n independent success/failure-type experiments, each of which yields success with a probability (or fairness factor) p. The probability of getting exactly x successes in n trials for a specified fairness value, p, is

$$\boxed{P = \frac{n!}{x!\,(n-x)!}\, p^{x}\,(1 - p)^{n-x}}$$

Binomial distribution



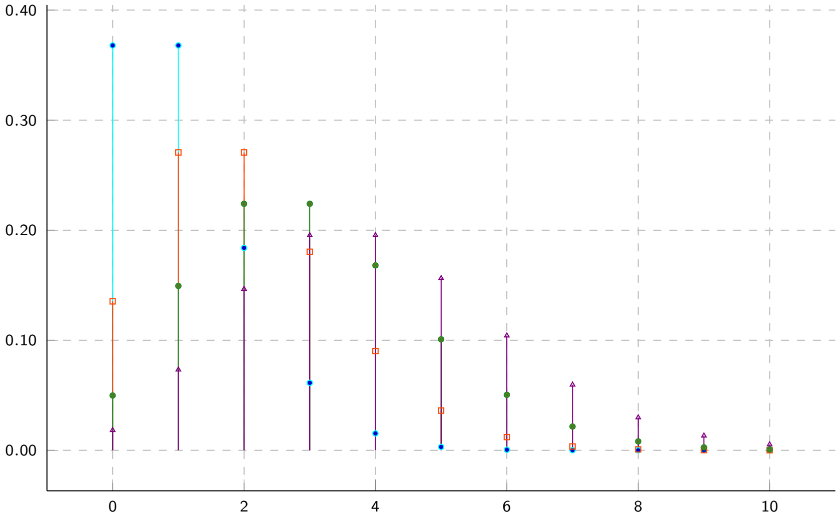
Poisson distribution _{Discrete}

Applicable when

a given number of events occur in a fixed interval of time if these events occur with a known average rate, independently of time since the last event, and two of them cannot occur at the same time. The probability of observing m events in an interval with the average number of events in an interval designated by λ is

$$\boxed{P\left(m\right) \;=\; \frac{\lambda^{m}\, e^{-\lambda}}{m!}}$$

Poisson distribution



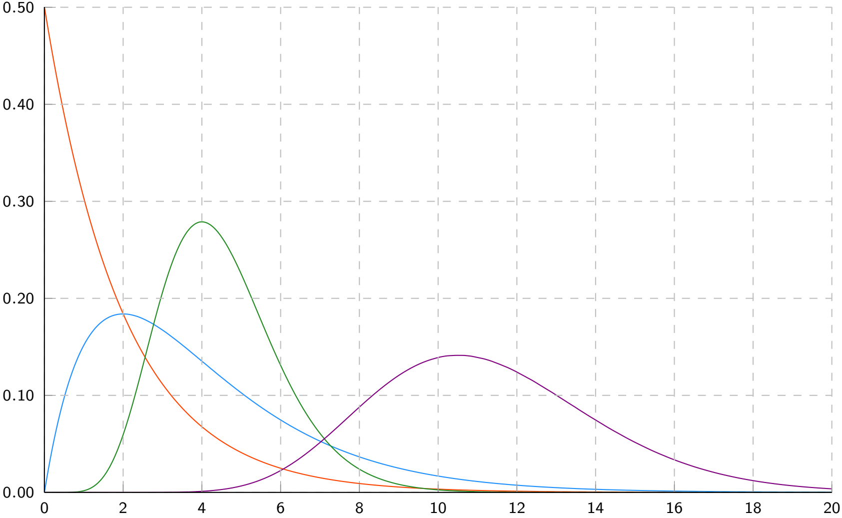
Gamma distribution

Applicable when

the waiting times between Poisson distributed events are relevant. The probability density function with shape parameter α and scale parameter β (inverse of rate parameter) is

$$\boxed{f(x) \;=\; \frac{1}{\Gamma\left(\alpha\right)\,\beta^{\alpha}}\,x^{\alpha-1}\,\exp\left(-\frac{x}{\beta}\right) \quad x \geqslant 0, \text{ and } \alpha, \beta > 0}$$

Gamma distribution



Normal/Gaussian distribution Continuous

Applicable as

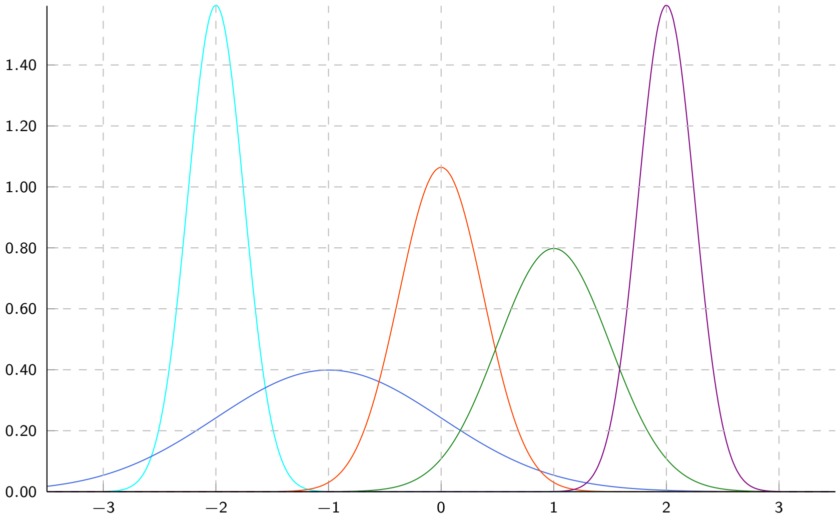
a limiting form of binomial distribution (De Moivre, 1733) and as a plausible distribution for measurement errors (Gauss, 1809). The probability density with mean μ and standard deviation σ is

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad -\infty < x, \mu < \infty, \text{ and } \sigma > 0}$$

Central limit theorem (Laplace)

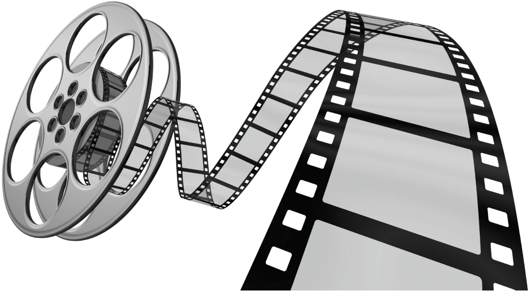
Under very general conditions when n random variables, whatever their distributions, are added together, the distribution of the sum tends towards the normal (i.e., bell shape) as n increases.

Normal/Gaussian distribution



Videos

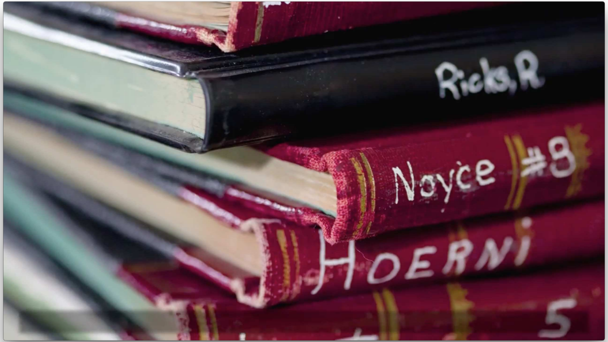
If a picture is worth a thousand words ...



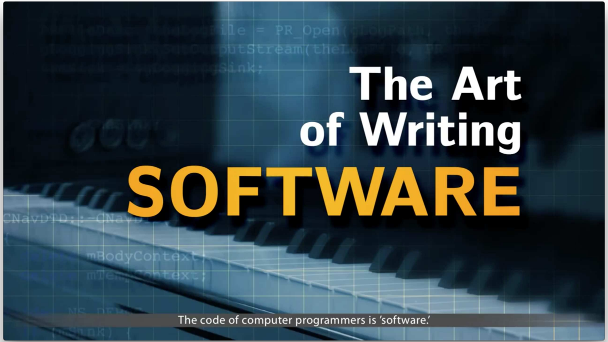
Computer History Museum

1401 N. Shoreline Blvd., Mountain View, CA 94043
(650) 810-1010

The Fairchild Notes



The Art of Writing Software



Supercomputing

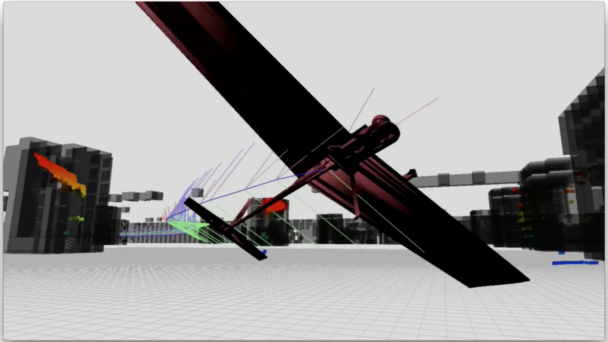
SC16

Salt Lake City, Utah | hpc matters.

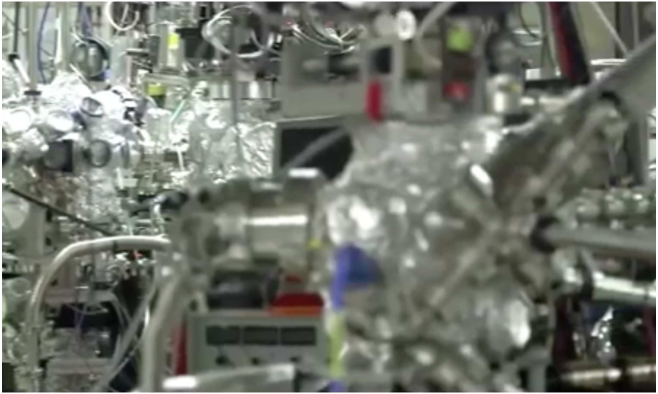
The International Conference for High Performance Computing, Networking, Storage and Analysis

An overview

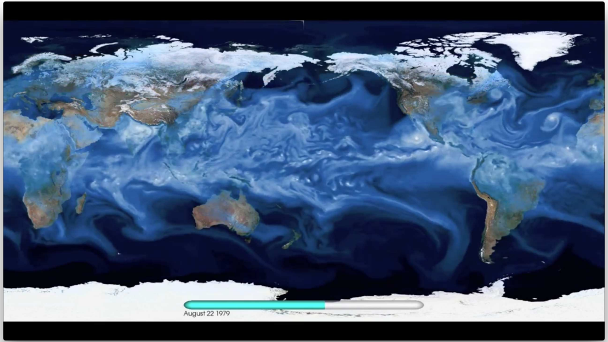
Aerospace



Batteries



Climate modeling



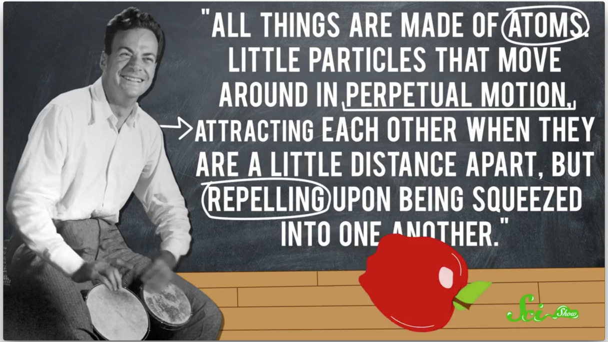
Diapers, detergents, shampoo

Entertainment

People and personalities

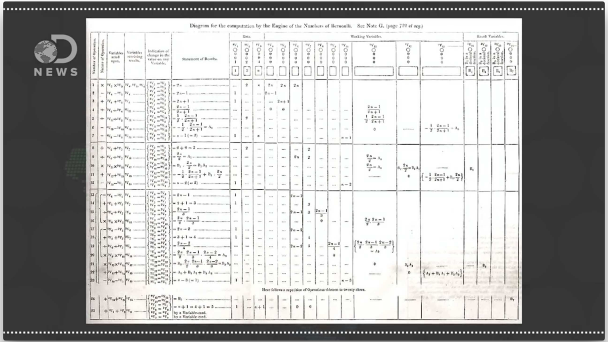
and their stories

Richard Phillips Feynman 1918 – 1988

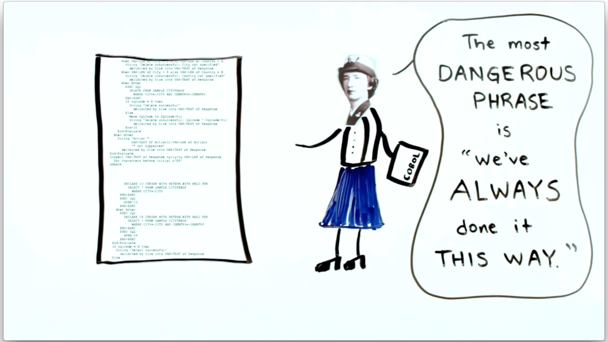


Alan Mathison Turing 1912 – 1954

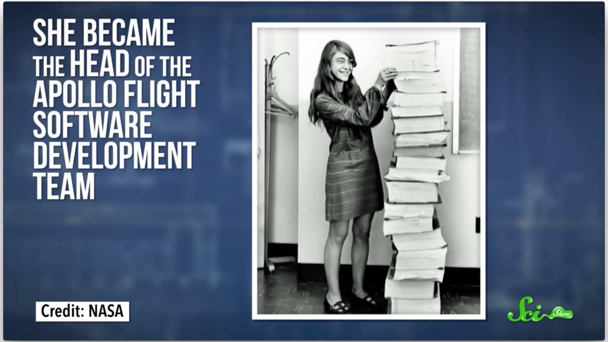
Ada Lovelace 1815 – 1852



Grace Brewster Murray Hopper 1906 – 1992



Margaret Heafield Hamilton 1936 – present

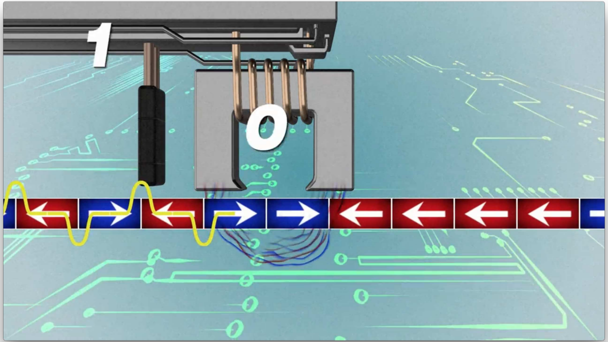


TED Ed

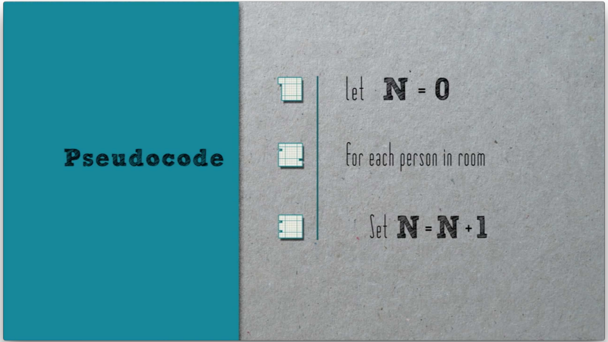
TED Ed

LESSONS WORTH SHARING

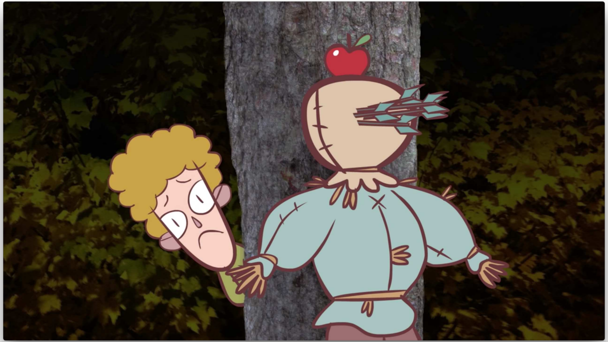
Hard Drives



Algorithm

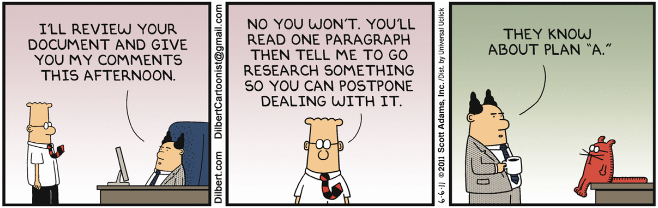


Accuracy vs Precision



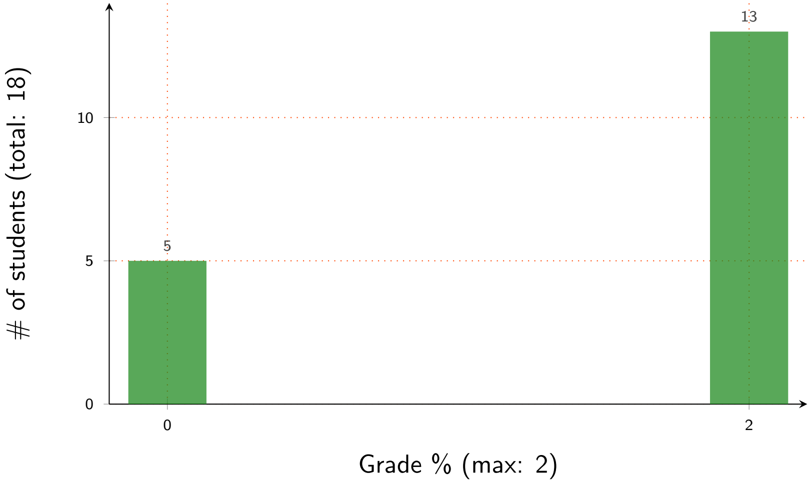
Review of Performance

How well have we been performing?

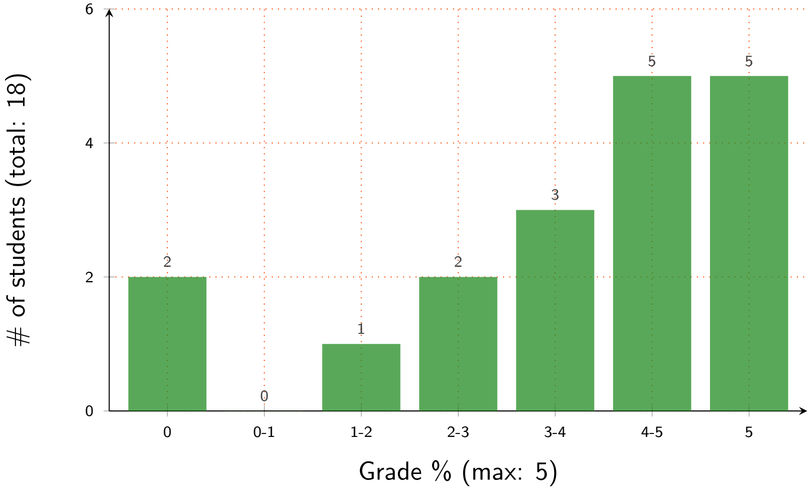


http://dilbert.com/strip/2011-06-06/

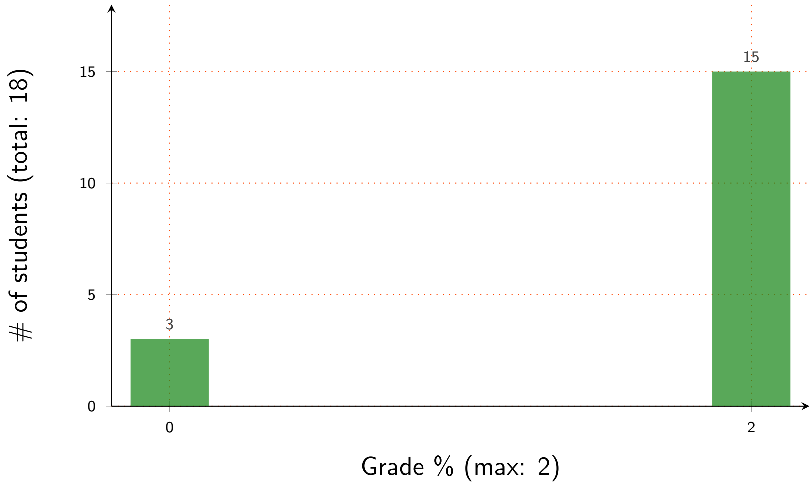
Active Participation #01 Research Marketing I: Twitter



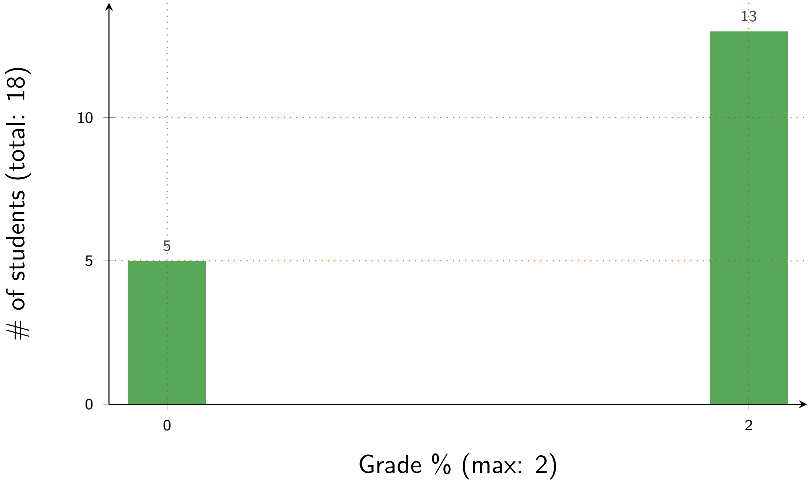
Assignment #01



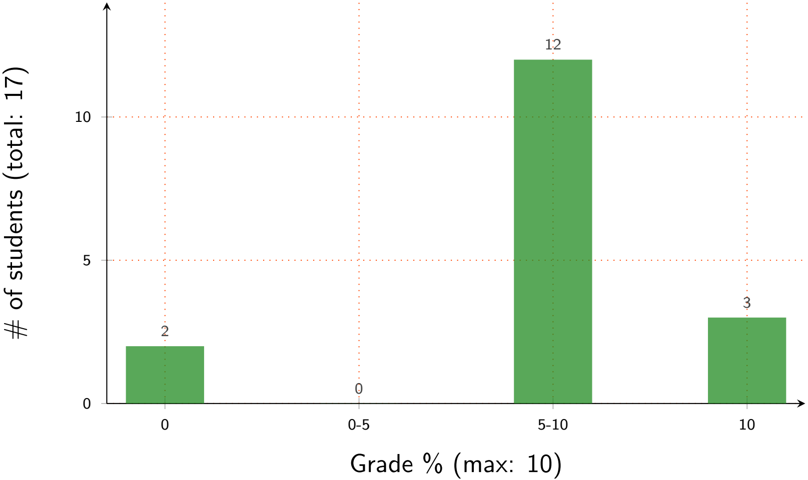
Active Participation #02 PB&J Sandwich Recipe



Active Participation #03 Research Marketing II: Professional/University Business Cards



Assignment #04



Superior and Top 500

A proposed compute node in Superior will have two Intel Xeon E5-2698 processors (each processor with 20 cores) at 2.20 GHz, 512 GB RAM, 480 GB Intel Enterprise SSD, Mellanox ConnectX-3 56 Gbps InfiniBand network, and will cost $13,263.13.

Ignoring the cost of physical space, racks, network, storage, electricity and labor, estimate the cost to build a #500 supercomputer (∼405 TFLOPS) with homogeneous compute nodes as the ones described above.

For a computer with N identical/homogeneous processors,

$$\text{FLOPS} = \text{N} \times \text{CPU speed} \times \frac{\text{FLOPs}}{\text{CPU cycle}}$$

Celsius $\longleftrightarrow$ Fahrenheit

Convert temperature between Celsius and Fahrenheit scales.

Is there a well-known technique to verify the conversion scheme?

Matrix elements

How many elements in a square matrix of order N? How will this number change if the matrix is upper (or lower) triangular?

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \qquad \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{nn} \end{pmatrix}$$

The impact and limitations of Moore's Law

Assuming that Moore's Law holds true, what is the speed up of a computer observed over an average adult's life in the US?

You and The Board Week #06: Ashley Kern

Drawing queens

Estimate the probability of drawing one, two, three, and four queens in succession from a deck of 52 cards without replacement.

Got questions?

If you do, find a way to contact me; and do so sooner than later

EERC B39 · (906) 487-4096 · g@mtu.edu · @sgowtham