COMPUTATIONAL APPROACHES

IN MATHEMATICAL ECOLOGY

Natalia Petrovskaya

School of Mathematics,

University of Birmingham, UK

The outline of the course (3 lectures, 1.5 hours each)

WE WILL DISCUSS:

- definition of basic numerical methods
- how to apply them in ecological problems

WE WILL NOT DISCUSS:

- programming techniques

http://web.mat.bham.ac.uk/N.B.Petrovskaya/NBPetrovskaya_teaching.htm

- software for ecological applications

R.L.Burden, J.D.Faires. *Numerical Analysis.* Brooks/Cole, CA, 2005

- basic mathematics behind numerical methods

The outline of the course (3 lectures, 1.5 hours each)

INTRODUCTION:

- Computational ecology beyond statistics
- Error analysis

HANDLING FUNCTIONS:

- Interpolation
- Numerical integration
- Finding roots

HANDLING FUNCTION DERIVATIVES:

- Numerical solution of ODEs
- Numerical solution of PDEs

LECTURE 1: Error analysis and function approximation

Why computational methods in ecology (apart from processing big data sets)?

- Complexity of ecological problems

$\downarrow$

- Complex mathematical models

$\downarrow$

- Solution in closed form is not available

$\downarrow$

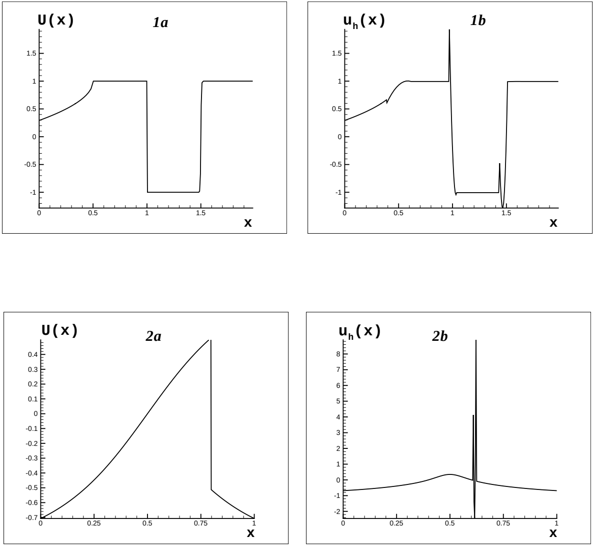
- Numerical solution

$\downarrow$

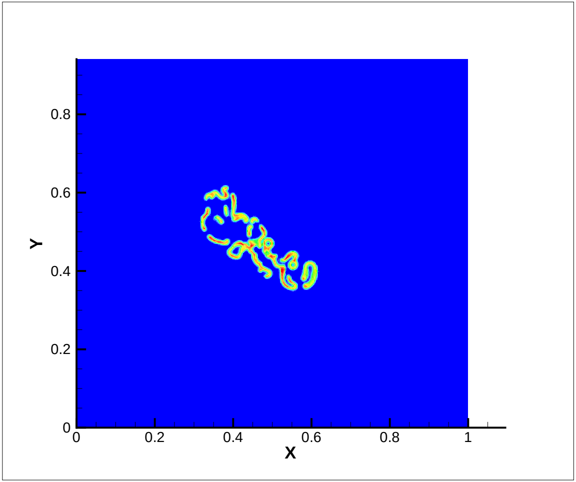
Is a numerical solution good enough? $\rightarrow$

reliable and accurate computational methods

What is wrong with the numerical solution?



How do we know that the numerical solution is correct?



Error analysis

- A wrong ecological hypothesis
- Errors in the mathematical model
- Measurement errors
- Truncation errors
- Round-off errors

Truncation error

- A truncation error is a characteristics of a numerical method used in the problem.

- Example: Consider the population size $n(t)$ and let the population growth rate be $dn(t)/dt$.
  - Replace $dn(t)/dt$ by a finite difference $(n_2 - n_1)/\delta t$, where n_2 and n_1 are the total number of a given species at time t and $t + \delta t$.
  - Assume zero error of the measurements.
  - The error of the method (as we replace the true derivative with a finite difference) has nothing to do with the measurement error.

Truncation error in ecological problems

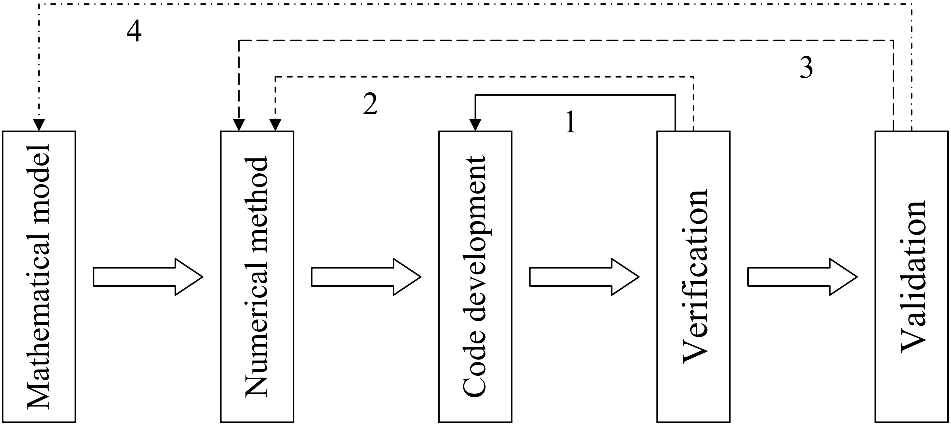
- In ecological problems, while a huge body of the research has been provided on the measurement errors, the truncation error related to the method has not been studied in detail.

- Example: Most of the sampling protocols currently used for the pest control imply that the truncation error is much smaller that the measurement error.
  - The theory states the truncation error is fully controllable and therefore the inherent error is of the utmost importance.
  - THIS IS NOT ALWAYS TRUE! (e.g. highly aggregated density distributions)
  - For a small number of samples the truncation error becomes a random error.

Truncation error

- A newly developed method is worthless without an error analysis!
- It does not make sense to use methods which introduce errors with magnitudes larger than the effects to be measured or simulated.
- On the other hand, using a method with very high accuracy might be computationally too expensive to justify the gain in accuracy.
- Basic means of control: the quality (e.g. a polynomial degree) of function or/and function derivative approximation, the time step size, the grid step size

Obtaining reliable and accurate numerical solutions



Obtaining reliable and accurate numerical solutions

- Preparation: specification of objectives, geometry, initial and boundary conditions, and available benchmark information; selection of the numerical method.
- Verification: a process for assessing simulation numerical uncertainty. Robustness of the simulation results should be proved by comparing them with the known analytical properties of the model, e.g. with exact solutions.
- Validation: a process for assessing simulation modelling uncertainty by using benchmark experimental data.

Summary

- Exploiting any computer program requires good understanding of (a) the ecological problem, (b) the mathematical model, and (c) a numerical method used in the code.
- Error analysis is a must! Never skip preparation, verification and validation steps when you solve a problem numerically.
- "Do use others people software but if you cannot understand a numerical algorithm behind the software, then never use it." (F.S. Acton, 1990)

References

- F.S. Acton. *Numerical Methods That Work.* Washington, DC: Math. Assoc. Amer., 2nd edition 1990.
- R.L.Burden, J.D.Faires. *Numerical Analysis. Brooks/Cole, Belmont,CA, 2005.*
- J.D. Hoffman. *Numerical Methods for Engineers and Scientists.* CRC Press, 2nd edition, 2001.
- S.V.Petrovskii, N.B.Petrovskaya. *Computational Ecology as an Emerging Science.* J.R.Soc. Interface Focus, 2012, vol.2(2), pp.241-254.

Function interpolation

in ecological problems

Ecological problem: pest insect monitoring and control

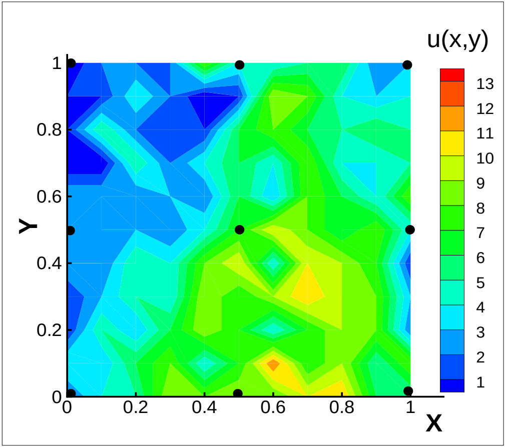
- The information about pest population size is obtained through trapping
- Once the samples (trap counts) are collected, the total number of the insects in the field is evaluated

The need in reliable methods to estimate the pest population size in order to avoid unjustified pesticides application and yet to prevent pest outbreaks.

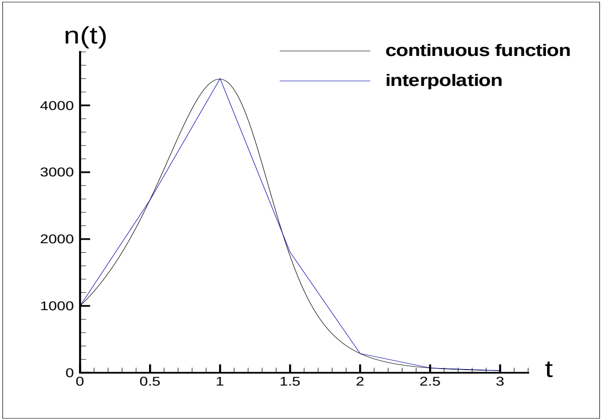
Experimental layout for pest insect monitoring

Example of spatial data: flatworm (*Arthurdendyus triangulatus*) spatial density distribution $u(x,y)$ reconstructed from field data.

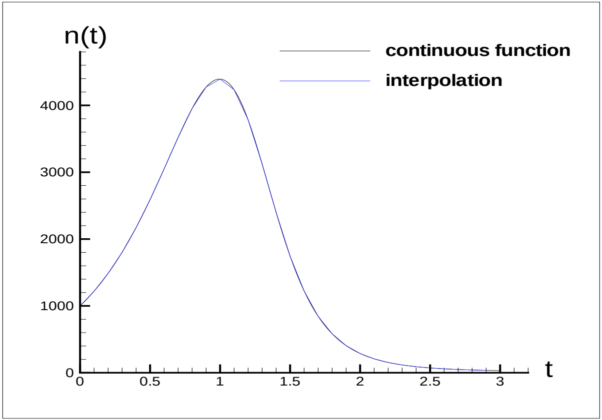
Given the pest density $u(x,y)$ at selected points, how can we reconstruct the pest insects density at any point (x,y)?



What is the accuracy of our evaluation?



How to achieve reliable accuracy?



References

- N.B.Petrovskaya, N.L.Embleton. *Computational Methods for Accurate Evaluation of Pest Insect Population Size.* W.A.C. Godoy and C.P. Ferreira (eds.), Ecological Modelling Applied to Entomology, Springer-Verlag Berlin Heidelberg, 2014.

- N.B.Petrovskaya, S.V.Petrovskii, A.K.Murchie. *Challenges of Ecological Monitoring: Estimating Population Abundance From Sparse Trap Counts.* J.R.Soc.Interface, 2012, vol.9(68), pp.420-435

- R.L.Metcalf, W.H.Luckmann (eds) *Introduction to Insect Pest Management.* Wiley, London, 1982.

- T.R.E. Southwood, P.A.Henderson PA *Ecological Methods.* Blackwell Science Ltd., Oxford, 2000.

An interpolation problem: outline

- Interpolation problem statement
- 1-d polynomial interpolation: Lagrange interpolation, interpolation by divided differences
- Interpolation error
- Piecewise polynomial interpolation
- 2-d polynomial interpolation

Interpolation problem

- *Given the pairs* $(\mathbf{x}_0, \mathbf{F}_0)$, $(\mathbf{x}_1, \mathbf{F}_1)$,..., $(\mathbf{x}_N, \mathbf{F}_N)$*, the problem of interpolation is to find an approximate value of* $f(\mathbf{x})$ *that corresponds to any selected value of* $\mathbf{x} \in D$*.*

- 1-d interpolation: Consider a 1-d domain $[a, b]$. Let only one value F_i be defined at each point x_i and $F_i \equiv f_i$. For such data the interpolation problem can be formulated as: *Given the pairs* (x_0, f_0), (x_1, f_1),... (x_N, f_N)*, and an arbitrary point* $x \in [a, b]$*, find an approximate value of* $f(x)$*.*

- Straightforward solution: *replace* $f(x)$ *with a polynomial* $p(x)$.

Polynomial interpolation

- How do we know that the polynomial interpolation is good enough for our problem?
- *The Weierstrass approximation theorem.*

- How can we construct a polynomial $p(x)$ that will interpolate $f(x)$?
- *We have to use all input information given to us.*

Polynomial interpolation

- We have $N+1$ pairs (x_i, f_i), $i = 0, 1, 2 \ldots, N$:

$$f(x) \approx p(x) = \sum_{k=0}^{N} a_k \phi_k(x),$$

where $\phi_k(x)$ are polynomial basis functions chosen for the approximation.

- Fit the polynomial to data:

$$p(x_i) = f(x_i), \quad i = 0, 1 \ldots, N.$$

- Solve for unknown coefficients $\mathbf{a} = (a_0, a_1, \ldots, a_N)$,

$$\mathbf{Va} = \mathbf{f}.$$

Monomial basis

$$f(x) \approx p(x) = \sum_{k=0}^{N} a_k x^k.$$

$$a_0 + a_1 x_0 + \ldots + a_N x_0^N = f_0,$$

$$a_0 + a_1 x_1 + \ldots + a_N x_1^N = f_1,$$

$$\vdots$$

$$a_0 + a_1 x_N + \ldots + a_N x_N^N = f_N.$$

Because points $(x_0, x_1, \ldots, x_N)$ are distinct, the matrix inverse $\mathbf{V}^{-1}$ exists,

$$\mathbf{a} = \mathbf{V}^{-1}\mathbf{f}.$$

Lagrange interpolation

$$\phi_k(x) \equiv L_k(x) = \prod_{l \neq k} \frac{x - x_l}{x_k - x_l}, \quad k = 0, 1, \ldots, N.$$

$$L_k(x_i) = \delta_{ik} = \begin{cases} 1, & \text{if } i = k, \\ 0, & \text{if } i \neq k. \end{cases}$$

$$p(x) = \sum_{k=0}^{N} f(x_k) \prod_{l \neq k} \frac{x - x_l}{x_k - x_l}$$

Interpolation by divided differences

For *any* function g(x) the divided differences are

$$g(x_i, x_j) = (g(x_i) - g(x_j))/(x_i - x_j),$$
$$g(x_i, x_j, x_k) = (g(x_i, x_j) - g(x_j, x_k))/(x_i - x_k),$$
$$\vdots$$
$$g(x_i, x_j \ldots, x_m, x_p) = (g(x_i, \ldots, x_m) - g(x_j, \ldots, x_p))/(x_i - x_p)$$

Any polynomial is $p(x) = p(x_0) + (x - x_0)p(x_0, x_1) + \ldots (x - x_0)(x - x_1) \ldots (x - x_{N-1})p(x_0, x_1, \ldots, x_N)$

We have $f(x) \approx p(x), \quad p(x_k) = f(x_k)$:

$$f(x) \approx f(x_0) + \sum_{k=1}^{N} a_k \phi_k(x) =$$

$$f(x_0) + \sum_{k=1}^{N} f(x_0, x_1, \ldots, x_k)(x - x_0)(x - x_1) \ldots (x - x_{k-1}).$$

Interpolation error

- The accuracy of interpolation depends on
  - the total length of the interval $[a, b]$, where the points x_i are located,
  - the polynomial degree N that we use for the interpolation.

- The interpolation error $E(x)$ at the point x

$$E(x) = |f(x) - p(x)|$$

How to estimate the interpolation error $E(x)$ if $f(x)$ is not available?

$$E(x) = |f(x) - p(x)| = \Big|\frac{f^{(N+1)}(s)h^{N+1}}{(N+1)!}\Big|,$$

$$E(x) < Ch^{N+1}.$$

Interpolation error

x_0	x_1	$p_1(1)$	$e(1)$	$e_r(1)$
0	2	8.66025	6.16025	2.46410
0.5	1.5	4.13924	1.63924	0.655695
0.75	1.25	2.91612	0.416123	0.166449
0.875	1.125	2.60443	0.104427	0.0417709

Example: Linear interpolation of the function $f(x) = 5x^2 sin(\frac{\pi}{6}x)$ at the point $x = 1$.

Piecewise interpolation

- Example: Consider $(x_0 = a, x_1, x_2, x_3 = b)$ and let $f_i = f(x_i)$, $i = 0, 1, 2, 3$.

- We can construct an cubic polynomial $p(x) = \sum_{k=0}^{3} a_k x^k$, $p(x_i) = f(x_i)$, $i = 0, \ldots, 3$.

- Alternatively, piecewise linear interpolation $p_m(x) = c_0^m + c_1^m x$, where $x \in [x_m, x_{m+1}]$, $m = 0, 1, 2$.

$2 - D$ interpolation

- Let the function $f(x, y)$ be defined at nodes (x_i, y_j) of a rectangular grid, $f_{ij} \equiv f(x_i, y_j)$.

- Given the values f_{ij}, the problem of interpolation is to find an approximate value of $f(x, y)$ corresponding to any selected point $(x, y) \in D$.

- We can interpolate $f(x, y)$ by applying consequent interpolation to each coordinate x and y.

$2-D$ interpolation

- Consider $1-D$ interpolation in the x-direction for any fixed $j = 0, 1, 2, \ldots, N_2$:

$$\tilde{f}_j(x) = \sum_{i=0}^{N_1} f_{ij} \prod_{p \neq i}^{N_1} \frac{x - x_p}{x_i - x_p}$$

- Given the values $\tilde{f}_j(x)$, consider $1-D$ interpolation in the y-direction:

$$p(x, y) = \sum_{j=0}^{N_2} \tilde{f}_j(x) \prod_{q \neq j}^{N_2} \frac{y - y_q}{y_j - y_q}$$

- The resulting interpolation formula is

$$f(x, y) \approx p(x, y) = \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} f_{ij} \prod_{p \neq i}^{N_1} \prod_{q \neq j}^{N_2} \frac{x - x_p}{x_i - x_p} \frac{y - y_q}{y_j - y_q}$$

Interpolation methods: checklist (incomplete!)

- Are you going to interpolate a function by polynomials?
- How much data are available to you? For a cloud of points it may be better to use LS approximation. For sparse data the accuracy may not be as expected.
- Check what data are available to you. Can you use a standard interpolation algorithm?
- Decide whether you want to use interpolation by a single polynomial or piecewise interpolation.

References

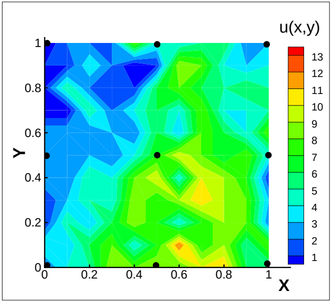
- P.J.Davis. *Interpolation and Approximation.* Dover Pub.Inc. 1975.
- W.E.Grove. *Brief Numerical Methods.* Englewood Cliffs, N.J. : Prentice-Hall, 1966.
- E.Isaacson and H. B. Keller. *Analysis of Numerical Methods.* New York ; London : Wiley, 1966.
- W.H.Press, S.A. Teukolsky, W.T. Vetterling, B.P.Flannery. *Numerical Recipes: The Art of Scientific Computing (3rd ed.).* New York: Cambridge University Press, 2007.

Methods of numerical integration

in ecological problems

Example of spatial data: flatworm (*Arthurdendyus triangulatus*) spatial density distribution $u(x, y)$ reconstructed from field data

The trap counts in the domain D are converted into the values $u_i \equiv u(x_i, y_i)$ of the pest insect population density $u(x, y)$ at locations $\mathbf{r}_i = (x_i, y_i)$, $i = 1, \ldots, N$.



Numerical integration in the pest control problem

- If the density $u(x, y)$ is known at any point (x, y) of the domain D, the total pest population size I is given by

$$I = \iint\limits_D u(x, y) dxdy.$$

- For given *precise* values $u_i \equiv u(x_i, y_i)$, $i = 1, \dots, N$, the pest population size I is reduced to computation of a weighted sum of the values u_i,

$$I \approx I_a(N) = \sum_{i=1}^{N} \omega_i u_i.$$

- The *approximation error* (integration error) depends on N,

$$e(N) = \frac{|I - I_a(N)|}{|I|}.$$

Numerical integration technique

- Generate a regular grid of N nodes in the unit square.
- Consider the values u_i, $i = 1, 2, \ldots, N$ at grid nodes.
- Replace $u(x, y)$ with polynomial $p_K^n(x, y)$ of degree K in the neighbourhood of node n.
- Integrate $p_K^n(x, y)$ instead of $f(x, y)$ (Newton-Cotes formulas)

$1-D$ numerical integration: Newton-Cotes formulas

- Let $x_1 = a$, $x_{i+1} = x_i + h$, $h = (b-a)/N$ and $f_i \equiv f(x_i)$
- Consider a polynomial $p_k(x)$ of degree k, $p_k(x_n) = f(x_n)$ for $n = i, i+1, \ldots, i+k$.

$$\int_{x_i}^{x_{i+k}} f(x)dx \approx a_i = \int_{x_i}^{x_{i+k}} p_k(x)dx$$

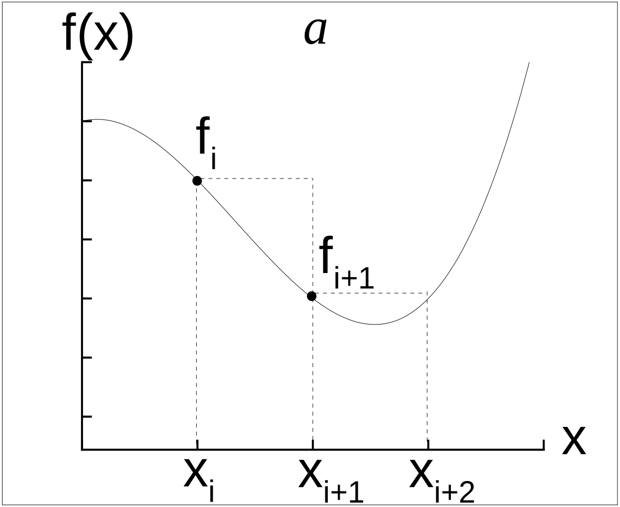
- (a) $p_0(x) = const$:

$$a_i = hf_i,$$

$$I \approx I_a(N) = \sum_{i=1}^{N} a_i = \sum_{i=1}^{N} hf_i = \sum_{i=1}^{N} \omega_i f_i$$

$\omega_i = h$, $i = 1, \ldots, N$

$1-D$ numerical integration: Newton-Cotes formulas



$1-D$ numerical integration: Newton-Cotes formulas

- (b) $p_1(x) = c_0 + c_1 x$:

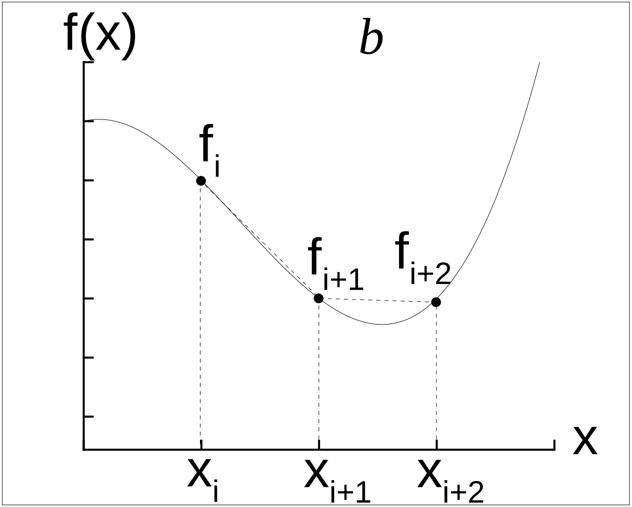
$$a_i = \frac{1}{2} h (f_i + f_{i+1}),$$

- Trapezoidal rule of integration

$$I \approx I_a(N) = \sum_{i=1}^{N} a_i = \frac{h}{2} \left[f_1 + 2 \sum_{i=2}^{N} f_i + f_{N+1} \right]$$

$\omega_1 = \omega_{N+1} = h/2$ and $\omega_i = h,\, i = 2, \ldots, N$

The trapezoidal rule of integration



$1-D$ numerical integration: Newton-Cotes formulas

- (c) $p_2(x) = c_0 + c_1 x + c_2 x^2$:

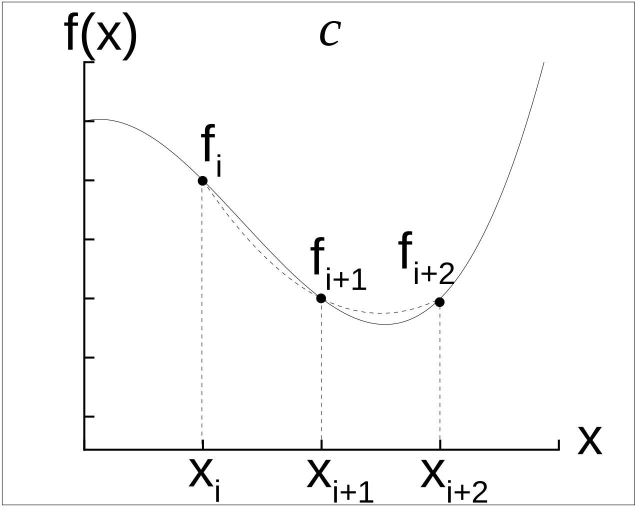
$$I_i = \int_{x_i}^{x_{i+2}} f(x)dx \approx \; a_i = \frac{1}{3}h(f_i + 4f_{i+1} + f_{i+2})$$

- Simpson's rule of integration

$$\int_a^b f(x)dx \approx I_a(N) = \frac{h}{3}\left[f_1 + 2\sum_{i=1}^{N/2-1} f_{2i+1} + 4\sum_{i=1}^{N/2} f_{2i} + f_{N+1}\right]$$

$$\omega_i = \frac{4h}{3},\; i = 2, 4, \ldots, N-1, \quad \omega_i = \frac{2h}{3},\; i = 3, 5, \ldots, N-2,$$
$$\omega_i = \frac{h}{3},\; i = 1, \text{ or } i = N$$

The Simpson rule of integration



Approximation (integration) error

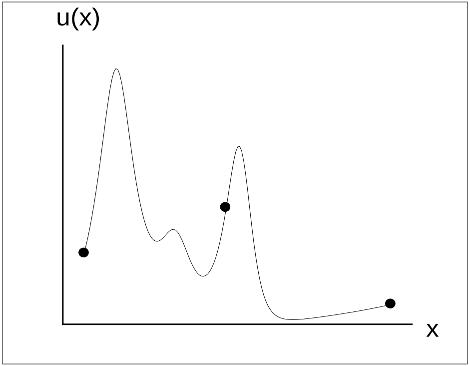
$$\text{Pest abundance } I \approx I_a(N) = \sum_{i=1}^{N} \omega_i u_i.$$

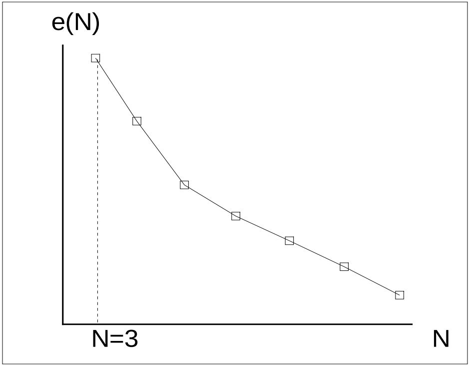
$$I_a(N) \to I, \text{ as } N \to \infty$$

$$e(N) = \frac{|I - I_a(N)|}{|I|} \to 0, \text{ as } N \to \infty.$$

- Given weight coefficients $\omega_i,\ i = 1, \ldots, N$, the approximation error depends on the number N of points where the data are available.
- For any fixed N the approximation error depends on a spatial pattern of the density function.

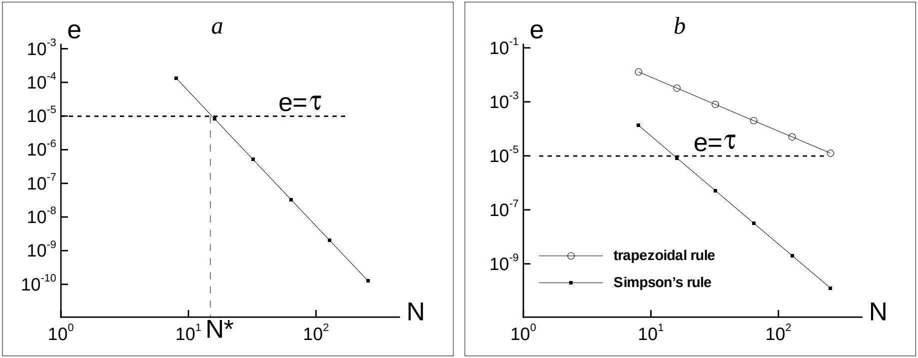
Example: approximation error for a $1 - d$ density function





Convergence and tolerance

- convergence rate $e = O(h^p)$ (asymptotic convergence estimate)
- tolerance $\tau : e \leq \tau$



$2 - D$ integration in rectangular domains

- Let the function $f(x, y)$ be defined at nodes (x_i, y_j) of a rectangular grid, $f_{ij} \equiv f(x_i, y_j)$.
-
$$I = \int_0^1 \int_0^1 f(x, y) dx dy = \sum_{i,j} I_{ij},$$

  where

$$I_{ij} = \int_{x_i}^{x_{i+1}} \int_{y_j}^{y_{j+1}} f(x, y) dx dy.$$

- The integration problem is reduced to the integral evaluation in each sub-domain $c_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$

$2-D$ integration in rectangular domains

- Consider a $1-D$ integral

$$I_{ij} = \int_{y_j}^{y_{j+1}} F(y)dy,$$

where

$$F(y) = \int_{x_i}^{x_{i+1}} f(x,y)dx.$$

- Employ $1-D$ Newton-Cotes formulas in order to evaluate the function $F(y)$

- Example: Trapezoidal rule of integration

$$I_{ij} \approx \frac{h^2}{4}\left[f(x_i,y_j)+f(x_{i+1},y_j)+f(x_i,y_{j+1})+f(x_{i+1},y_{j+1})\right]$$

Numerical integration for pest insect monitoring

Accuracy requirements are not very demanding:

$$e(N) \leq \tau,$$

where $\tau \sim 0.2 - 0.5$ is a specified tolerance.
N is small in field measurements $\Rightarrow$ (often but not always) inaccurate evaluation I_a of the pest abundance I.

- For any fixed (small) N the approximation error depends on a spatial pattern of the density function.

What is the number N of traps to provide the accuracy required in ecological applications?

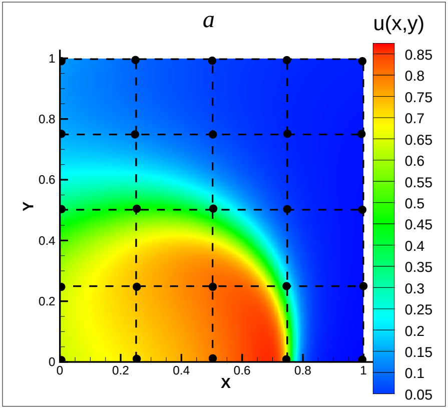
What accuracy can we expect when N is fixed?

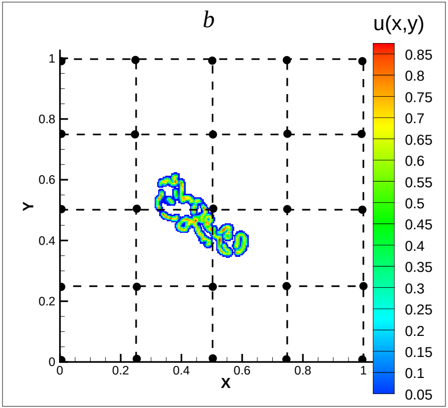
Can we rely upon convergence estimates $e = O(h^p)$?

Approximation error for different spatial density patterns

(a) $e(N) \sim 10^{-2}$

(b) $e(N) \sim 1.0$





Current challenges: "the coarse grid problem"

- Financial and labor resources available for monitoring are always limited.
- N is small in field measurements $\Rightarrow$ (in some cases) inaccurate evaluation I_a of the pest abundance I.
- Recognition of spatial patterns is extremely important!

References

- R.L.Burden, J.D.Faires. *Numerical Analysis.* Brooks/Cole, Belmont,CA, 2005.
- P.J.Davis P.Rabinowitz. *Methods of Numerical Integration.* Academic Press, New York, 1975.
- H.Engels. *Numerical Quadrature and Cubature.* Academic Press, New York, 1980.
- N.B.Petrovskaya, N.L.Embleton. *Computational Methods for Accurate Evaluation of Pest Insect Population Size.* in W.A.C. Godoy and C.P. Ferreira (eds.), Ecological Modelling Applied to Entomology, Springer, 2014.

Basic concepts

- Approximation $f(x) \approx p(x)$
- Accuracy $E(x) \to 0$
- Convergence $E(x) = O(h^m)$
- Efficiency/convenience