Problem Set 4: Sorting, Root Finding, and Minimization
Computational Physics
Physics 480/680
James Sethna; Due Monday, March 17
Last correction at March 2, 2014, 8:21 am

Reading
Numerical Recipes chapters 8, 9, and 10, skimming the technical bits

4.1 **The Birthday Problem.** (Sorting, Random numbers) ②

Remember birthday parties in your elementary school? Remember those years when two kids had the same birthday? How unlikely!

How many kids would you need in class to get, more than half of the time, at least two with the same birthday?

(a) Numerical. *Write a routine* `BirthdayCoincidences(K, C)` *that returns the fraction among C classes for which at least two kids (among K kids per class) have the same birthday.* (Hint: By sorting a random list of integers, common birthdays will be adjacent.) *Plot this probability versus K for a reasonably large value of C. Is it a surprise that your classes had overlapping birthdays when you were young?*

One can intuitively understand this, by remembering that to avoid a coincidence there are $K(K-1)/2$ pairs of kids, all of whom must have different birthdays (probability $364/365 = 1 - 1/D$, with D days per year).

$$P(K, D) \approx (1 - 1/D)^{K(K-1)/2} \tag{1}$$

This is clearly a crude approximation – it doesn't vanish if $K > D$! Ignoring subtle correlations, though, it gives us a net probability

$$\begin{aligned} P(K, D) &\approx \exp(-1/D)^{K(K-1)/2} \\ &\approx \exp(-K^2/(2D)) \end{aligned} \tag{2}$$

Here we've used the fact that $1 - \epsilon \approx \exp(-\epsilon)$, and assumed that K/D is small.

(b) Analytical. *Write the exact formula giving the probability, for K random integers among D choices, that no two kids have the same birthday.* (Hint: What is the probability that the second kid has a different birthday from the first? The third kid has a different birthday from the first two?) *Show that your formula does give zero if $K > D$. Converting the terms in your product to exponentials as we did above, show that your answer is consistent with the simple formula above, if $K \ll D$. Inverting equation 2, give a formula for the number of kids needed to have a 50% chance of a shared birthday.*

Some years ago, we were doing a large simulation, involving sorting a lattice of 1000^3 random fields (roughly, to figure out which site on the lattice would trigger first). If we want to make sure that our code is unbiased, we want different random fields on each lattice site – a giant birthday problem.

Old-style random number generators generated a random integer (2^{32} 'days in the year') and then divided by the maximum possible integer to get a random number between zero and one. Modern random number generators generate all 2^{52} possible doubles between zero and one.

(c) *If there are* $2^{31} - 1$ *distinct four-byte positive integers, how many random numbers would one have to generate before one would expect coincidences half the time? Generate lists of that length, and check your assertion.* (Hints: It is faster to use array operations, especially in interpreted languages. I generated a random array with N entries, sorted it, subtracted the first $N - 1$ entries from the last $N - 1$, and then called *min* on the array.) *Will we have to worry about coincidences with an old-style random number generator? How large a lattice* $L \times L \times L$ *of random double precision numbers can one generate with modern generators before having a 50% chance of a coincidence? If you have a fast machine with a large memory, you might test this too.*

4.2 **Washboard Potential.** (Solving) ②

Consider a washboard potential[1]

$$V(r) = A_1 \cos(r) + A_2 \cos(2r) - Fr \tag{3}$$

with $A_1 = 5$, $A_2 = 1$, and F initially equal to 1.5.

(a) *Plot* $V(r)$ *over* $(-10, 10)$. *Numerically find the local maximum of* V *near zero, and the local minimum of* V *to the left (negative side) of zero. What is the potential energy barrier for moving from one well to the next in this potential?*

Usually finding the minimum is only a first step – one wants to explore how the minimum moves and disappears…

(b) *Increasing the external tilting field* F, *graphically roughly locate the field* F_c *where the barrier disappears, and the location* r_c *at this field where the potential minimum and maximum merge.* (This is a saddle-node bifurction.) *Give the criterion on the first derivative and the second derivative of* $V(r)$ *at* F_c *and* r_c. *Using these two equations, numerically use a root-finding routine to locate the saddle-node bifurcation* F_c *and* r_c.

[1] A washboard is what people used to hand-wash clothing. It is held at an angle, and has a series of corrugated ridges; one holds the board at an angle and rubs the wet clothing on it. Washboard potentials arise in the theory of superconducting Josephson junctions, in the motion of defects in crystals, and in many other contexts.

4.3 **Sloppy Minimization.** (Minimization) ③

"With four parameters I can fit an elephant. With five I can make it waggle it's trunk." This statement, attributed to many different sources (from Carl Friedrich Gauss to Fermi), reflects the problems found in fitting multiparameter models to data. One almost universal problem is *sloppiness* – the parameters in the model are poorly constrained by the data.[2]

Consider the classic ill-conditioned problem of fitting exponentials to radioactive decay data. If you know that at $t = 0$ there are equal quantities of N radioactive materials with half-lives γ_n, the radioactivity that you would measure is

$$y(t) = \sum_{n=0}^{N-1} \gamma_n \exp(-\gamma_n t). \tag{4}$$

Now, suppose you don't know the decay rates γ_n. Can you reconstruct them by fitting the data to experimental data $y_0(t)$?

Let's consider the problem with two radioactive decay elements $N = 2$. Suppose the actual decay constants for $y_0(t)$ are $\alpha_n = n+2$ (so the experiment has $\gamma_0 = \alpha_0 = 2$ and $\gamma_1 = \alpha_1 = 3$), and we try to minimize the least-squared error[3] in the fits C:

$$C[\boldsymbol{\gamma}] = \int_0^\infty (y(t) - y_0(t))^2 \, dt. \tag{5}$$

You can convince yourself that the least-squared error is

$$\begin{aligned} C[\boldsymbol{\gamma}] &= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \Big(\frac{\gamma_n \gamma_m}{\gamma_n + \gamma_m} \\ &\qquad\qquad + \frac{\alpha_n \alpha_m}{\alpha_n + \alpha_m} - 2\frac{\gamma_n \alpha_m}{\gamma_n + \alpha_m} \Big) \\ &= \frac{49}{10} + \frac{\gamma_0}{2} - \frac{4\,\gamma_0}{2+\gamma_0} - \frac{6\,\gamma_0}{3+\gamma_0} \\ &\quad + \frac{\gamma_1}{2} - \frac{4\,\gamma_1}{2+\gamma_1} - \frac{6\,\gamma_1}{3+\gamma_1} + \frac{2\,\gamma_0\,\gamma_1}{\gamma_0+\gamma_1}. \end{aligned}$$

(a) *Draw a contour plot of C in the square $1.5 < \gamma_n < 3.5$, with enough contours (perhaps non-equally spaced) so that the two minima can be distinguished. (You'll also need a fairly fine grid of points.)*

One can see from the contour plot that measuring the two rate constants separately would be a challenge. This is because the two exponentials have similar shapes, so

[2]'Elephantness' says that a wide range of behaviors can be exhibited by varying the parameters over small ranges. 'Sloppiness' says that a wide range of parameters can yield approximately the same behavior. Both reflect a *skewness* in the relation between parameters and model behavior.

[3]We're assuming perfect data at all times, with uniform error bars.

increasing one decay rate and decreasing the other can almost perfectly compensate for one another.

(b) *If we assume both elements decay with the same decay constant* $\gamma = \gamma_0 = \gamma_1$, *minimize the cost to find the optimum choice for* γ. *Where is this point on the contour plot? Plot* $y_0(t)$ *and* $y(t)$ *with this single-exponent best fit on the same graph, over* $0 < t < 2$. *Do you agree that it would be difficult to distinguish these two fits?*

This problem can become much more severe in higher dimensions. The banana-shaped ellipses in your contour plot can become needle-like, with aspect ratios of more than a thousand to one (about the same as a human hair). Following these thin paths down to the true minima can be a challenge for multidimensional minimization programs.

(c) *Find a method for storing and plotting the locations visited by your minimization routine (values of* (γ_0, γ_1) *at which it evaluates* C *while searching for the minimum). Starting from* $\gamma_0 = 1, \gamma_1 = 4$, *minimize the cost using as many methods as is convenient within your programming environment, such as Nelder-Mead, Powell, Newton, Quasi-Newton, conjugate gradient, Levenberg-Marquardt (nonlinear least squares)…. (Try to use at least one that does not demand derivatives of the function). Plot the evaluation points on top of the contour plot of the cost for each method. Compare the number of function evaluations needed for the different methods.*