In Search of Various Oh's

Find a and b such that: $f(n) = an^b - R(n)$

- $R(n) \geq 0$, but hopefully small, for all $n > n_0$
- $R(n) \leq 0$, but hopefully small, for all $n > n_0$
- $|R(n)|$ as small as possible, for all $n > n_0$

First things a statistician will probably want to talk about:

Where are the physical sources of variation?

- problem-to-problem for the same n ...
- computer-to-computer for the same problem ...
- execution-to-execution for the same computer ...

Where are the structural uncertainties that cannot be avoided?

- functional form of R ...
- possibility that a isn't really constant, even if $O(a) = 1$ ...
- possible "granular" response to discrete n (e.g. discontinuous R) ...

"**<u>Find $\mathbf{a_L}$, $\mathbf{b_L}$, $\mathbf{a_U}$, and $\mathbf{a_U}$</u>** such that

$$a_L n^{b_L} \; < \; f(n) \; < \; a_U n^{b_U} \; ...$$

(*what statisticians don't do much*) ... is provably true for all functions in a specified class, perhaps assuming a relationship between the observed n's and n_0."

(*what statisticians do more of*) ... is true except with some controllable and quantifiable risk* for functions in a perhaps richer class."

$*$ relative to the sources of variability, noise, and uncertainty previously mentioned

Standard regression methods ...

- are good for modeling the response near the data
- are generally not so good for revealing model structure

They typically produce confidence bounds that grow to asymptotic uselessness with n ... this will make them of little value here.

Generally need to add information/assumptions to reflect how structure is more apparent with larger n (same intuition as with PW3).

<u>Statistical intuition toward this end</u>: Need information concerning:

- $\underline{a}n^{\underline{b}}$ (2 degrees of freedom)
- How large is R relative to a?
- How quickly does R die out with n?
- How simple/smooth/crazy is R? (...min 5 d.f. so far)

If there is also rough/"discontinuous" (in n) noise

- How large, relative to a?
- How quickly does it die out?

Sounds like you need ... well, maybe I need ... substantially more than 5 data points. (Statisticians are famous for saying things like this.)

How about this?

$$
\begin{aligned}
f(n) &= an^b(\text{dominant}) + a_1 n^{b_1} + a_2 n^{b_2} + ... \\
&= an^b[1 + \tfrac{a_1}{a} n^{-(b-b_1)} + \tfrac{a_2}{a} n^{-(b-b_2)} + ...] \\
ln(f(n)) &= ln(a) + b \times ln(n) + ln[''] \\
&\approx ln(a) + b \times ln(n) + \{r_1 n^{-\delta_1} + r_2 n^{-\delta_2} + ...\}
\end{aligned}
$$

Model $Z(n) = \{-\}$ as a random function with:

- $E[Z(n)] = 0$
- $SD[Z(n)] = \sigma n^{-\delta}$ (size and decay rate of extra)
- $Corr[Z(n), Z(n')] = exp(-\theta[ln(n) - ln(n')]^2)$ (“smoothness”)

Think about lower and upper confidence limits for b ...

Relatively vague priors, design $= \{2, 4, 8, ..., 1023\}$, MCMC, 2.5% and 97.5% points of posterior:

function	$\hat{b}_L$	$\hat{b}_U$
$3n^{.2} + 100$ (#2)	0.169	0.175
$3n^{.8} - n^{.2}$ (#6)	0.822	0.853
$3n^{.8} + n^{.6}$ (#8)	0.834	0.848
$3n^{1.2} - 2n^{.8} + n^{.4}$	1.158	1.168

Excuses: In each case, $\hat{\delta}$ was very, very small ... suggesting that the model isn't tracking the "smaller-term decay" adequately.