When do we stop Newton's method?
What can we say about statistically
insignificant difference between 2 models?
$\delta\chi^2 \ll 1$ ⇒ I don't prefer one over the other

curvature = $A^T N^{-1} A$
param errors $\left\{ \mathrm{diag}\left[(A^T N^{-1} A)^{-1} \right] \right\}^{1/2}$

Newton's: each new step is (hopefully) a better model.
When should I quit? well if χ^2 improves by $\ll 1$
or new model isn't any better.

also, if parameter shifts $\ll$ errors on parameters
then we can also quit.

stop when χ^2 stops changing (relative to 1)
or parameters stop changing (relative to their errors)

How do we get data errors?

#1 choice: angels on high tell us the answer

#2 choice: if we don't have them, we need to figure out/guess at them.

in simple world, maybe errors are constant an uncorrelated. In this case, the cost variance of residual as an estimate.

decision points include: is $\sigma(i)$ constant?

is $\langle \sigma_i \sigma_j \rangle = 0$?

is noise a simple or complicated setup?

$$\overline{V} = \sum \frac{(y_i - m_i)^2}{\sigma_i^2} \Big/ N_{data} \qquad \text{normalized}$$

i/ne fit quality

$N_{data} \Rightarrow N_{DoF} = N_{data} - n_{params}$

do I add a parameter? Look at Bayesian evidence

$\langle (\bar{d} - d_{true})^2 \rangle$

we take many data points
w/ same mean. How close is
average to the mean? (Statistical)

$Var(d_i - d_{true}) = \sigma_i^2$

$Var(a+b) = Var(a) + Var(b)$ iff a & b are uncorrelated.

$Variance(Ca) = C^2 Var(a)$

$Var(\bar{d})? \Rightarrow \frac{\sum d_i}{n} \Rightarrow \frac{n\, Var(d_i)}{n^2}$

$\Rightarrow Var(\bar{d}) = \frac{Var(d)}{n}$

↑ got $Var(d)$ from look at variance within chain

$Var(\text{chain mean}) \neq Var(\text{with chain})$

↑ independent samples

$$Var(\bar{ch}_1 - \bar{ch}_2) \Rightarrow \frac{Var(ch_1)}{n_{samp_1}} + \frac{Var(ch_2)}{n_{samp_2}}$$

if chains are same length:

$$n_{samp1} = n_{samp2}$$

Importance Sampling

add new likelihood:

new phase space density = new - old phase space

old phase space is just our old chain

new phase = new . chain

(incremental)

⇒ we can weight chain samples

by their likelihood.

if chain didn't move too much, we might still be confused

Example: High T chain

for most σ error bars

$P_1/P_2 = e^{-\delta x^2/2T}$, but we use $T=1$

$T=$ large, we scale down δx^2

all of a sudden large δx^2 isn't

lets us probe wider σ, but

we need to handle correctly

we know in high T chain

$P_1 \propto e^{-\delta x^2/2T} \Rightarrow$ in truth $e^{-\delta x^2/2}$ $e^{-\delta x^2/2}$ / $e^{-\delta x^2/2T}$

$T=1$

$\Rightarrow e^{-\delta x^2/2}(1-\frac{1}{T})$

So I can take a high T chain

weight each sample by

$e^{-\delta x^2/2}(1-\frac{1}{T})$

to get chain we want.

Normally $\bar{x} = \frac{\sum x_i}{n} \Rightarrow \frac{\sum x_i w_i}{\sum w_i}$

Correlated noise realizations

Cholesky decomp PD matrix A

is $LL^T = A$, L is triangular