Long Range Dependence

Discussed 1/18/01

Richard Smith, Michele Trovero & Steve Marron

Data Setting A:

1 million consecutive packets

- time stamps & packet sizes
- both incoming and outgoing
- from UNC Main Link
- total time ~ 200 secs (~ 3 mins).

Data Setting A (cont):

Treat binned data at fixed scale:

- binwidth, i.e. “scale” $m \approx 0.02\,\text{sec}$
- ~10,000 bins
- ~ 100 obs’s per bin

Long Range Dependence Analysis, I

Autocovariance: of "binned" time series

Show UncLinkData2p12d1.ps

Observations:

- Have "mixture" of (~10%) long range dep. and (~90%) "white noise"?
- Related to Riedi & Willinger idea: "for $m >$round trip time" have Fractional Gaussian noise?

Long Range Dependence Analysis, I (cont.)

- study “exponential decay” via

$$y \approx a \cdot e^{b \cdot x} \Leftrightarrow \log y \approx b \cdot x + \log a$$

- exponential fit, $\phi \approx 0.989$, suggests “near unit root”, a type of “long range dependence”

- study “polynomial decay” via

$$y \approx a \cdot x^{b} \Leftrightarrow \log y \approx b \cdot \log x + \log a$$

- polynomial fit, with power = -0.16, suggests Hurst parameter $H \approx 0.92$ (strong “long range depend’ce”)

Long Range Dependence Analysis, II

Based on Periodograms (by Richard Smith)

Motivation: easier to “quantify uncertainty” in estimates

(e.g. no idea of “error” in above H estimate)

Show RichardSmithPlots1.ps and RichardSmithText1.ps

Result: estimated Fractional Differencing Parameter, d, with error bars.

Connection to above: $H \approx 0.92 \quad \Rightarrow \quad d = H - \frac{1}{2} \approx 0.42$, which is “within error bars”.

Big Questions:

Q1: Did we choose the “right scale”?

Q2: What happens “across scales”?

Data Setting B:

Same 1 million consecutive packets, but multi-scale view:

"Finest scale":

- binwidth, i.e. "scale" $m \approx 0.001\,\text{sec}$
- ~200,000 bins
- ~about 5 obs's per bin

$\vdots$

"Coarsest scale":

- binwidth, i.e. "scale" $m \approx 1\,\text{sec}$
- ~200 bins
- ~about 5000 obs's per bin

Long Range Dependence Analysis, I

Autocovariance: of "binned" time series

Show UncLinkData2p22d1.mpg

Coarser Scales $\Rightarrow$

- $\Rightarrow$ overall more dependence
- $\Rightarrow$ steeper at left
- $\Rightarrow$ more variability

Long Range Dependence Analysis, I (cont.)

Summaries of parameters:

Show UncLinkData2p22d1.ps

R^2 for long range dependence (% of SS “LRD”):

- “low” for “small” $m \in (10^{-3}, 10^{-2})$
- “increases” for $m \in (10^{-2}, 10^{-1})$
- “large” for $m \in (10^{-1}, 1)$

Long Range Dependence Analysis, I (cont.)

ϕ, Power, Hurst parameter:

- all three are closely correlated
- increasing for small scales $m \in \left(10^{-3}, 2*10^{-2}\right)$

Problem at large scales: too noisy (bin counts too sparse)

Long Range Dependence Analysis, II

(never performed for Setting B)

Data Setting C:

Idea: longer reach across scales
(requires more data, and different data storage)

New Multiple Scales (always 10,000 bins):

m(sec)	0.32	⋯	0.02	0.01	0.005	⋯	0.0003
$total$(sec)	3200	⋯	200	100	50	⋯	3.125

- longest ~ 1 hr, to avoid “time of day” effects
- kept length at 10,000 bins, to allow easy data-handling (and hope to “sufficiently dampen noise”)

Long Range Dependence Analysis, I

Multi-scale autocorrelation analysis:

Show UncLinkData4p1d1t2.mpg

Similar lessons,

but much lower large scale variability

Long Range Dependence Analysis, I (cont.)

Summaries of parameter estimates:

Show UncLinkData4p1d1t2.ps

L. R. D. R^2: similar to above, but stable for large m

ϕ, Power, Hurst parameter: similar to above, but now

- increasing for small scales $m \in (10^{-3}, 10^{-2})$

Long Range Dependence Analysis, II

(Michele Trovero)

i. Auto Correlation Function

Show Trovero\ln_size_acf.ps

- lessons similar to above

ii. Partial Correlation Function

Show Trovero\ln_size_pcf.ps

- Suggests AR (vs. ARMA) structure for all scales?

Long Range Dependence Analysis, II (cont.)

iii. Raw Periodogram

Show Trovero\ln_size_pdg.ps

- more “overall power” for larger m, since longer trace?
- Flat for small m, “tilted” for larger m

iv. Estimated Spectral density

Show Trovero\ln_size_sd.ps

- similar effects, but smoother

Long Range Dependence Analysis, II (cont.)

v. Estimates of Fractional ARIMA index d

Show Trovero\ln_size_d.ps

- near 0 for small m
- around 0.4 for large m, consistent with above
- errorbars seem too small?

Again show RichardSmithPlots1.ps, errorbars part

- Different samplings of “main curve”?