Long Range Dependence

Discussed 5/3/01

Steve Marron

Data Setting A:

1 million consecutive packets

- time stamps & packet sizes
- both incoming and outgoing
- from UNC Main Link
- gathered in 1998 (will look soon at newer data)
- total time ~ 200 secs (~ 3 mins).

Data Setting A (cont):

Study long range dependence:

1. Zooming SiZer analysis

Show UncLinkData2p8d1.mpg

- much “statist’ly significant” structure at coarse scales
- decreases for finer scales
- very few features even at smallest scale
- “long range dependence” artifacts?
- Don’t have this for Homogeneous Poisson Data

Show UncLinkSim1t1p3d1.mpg

Data Setting B:

Across scale analysis – 1998 data

New Multiple Scales (always 10,000 bins):

$m(\text{sec})$	0.32	$\cdots$	0.02	0.01	0.005	$\cdots$	0.0003
$total(\text{sec})$	3200	$\cdots$	200	100	50	$\cdots$	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 UncLinkData4p1d1t1.mpg

- smallest scales: ~ i.i.d.
- consistent with Bill Cleveland's analysis
- Long Range Dep. appears around $m = 10^{-3} = 0.001$
- "steady uplift" across scales ????
- Rolf Riedi Observation: difficult to track "time"

Long Range Dependence Analysis, I (cont.)

Rolf's suggestion 1: flag "time" in these plots

Show UncLinkData4p5d1t1.mpg

- also show broader range (100 lags $\rightarrow$ 2000 lags)
- blue line shows previous edge
- not really a "vertical uplift"
- but not "sweeping in with time" either

Long Range Dependence Analysis, I (cont.)

Rolf's suggestion 2: fix time scale

Show UncLinkData4p6d1t1.mpg

- blue line same as above (previous boundary)
- LRD happens first for smaller times????
- "Aggregation" moves it along to larger times????

Long Range Dependence Analysis, I (cont.)

Effect disappears for simulated Homogeneous Poisson data

Show UncLinkSim1t1p5d1.mpg

Possible new research:

Find point processes which have these properties:

- $M/G/\infty$ output
- Aggregated Cascaded On-Off Processes
- Melamed's TES
- Conservative Cascade modulated Poisson (Riedi)
- ARMA modulated Poisson (Davis)
- Cleveland's Weibull & dependent model

Big Questions:

Is this worthwhile?

Bill Cleveland: LRD is there, but not important. For useful traffic simulation, only need to study inter-arrival times

(and Poisson is good enough)

Caveat: Applies to “center of internet”, not “edges”

(Willinger showed queueing is affected by LRD “at edges”)

(OK for us, since these data live there)

Big Questions:

Bill's Main Reason: queueing properties of traffic only influenced by short time events.

- Is this correct? (what does it mean?)
- Only "first order approximation"?
- Is simul'n w/ grossly wrong large scale behavior OK?