Aalto University
School of Electrical Engineering

Network Traffic Measurements and Analysis

Lecture IV: Stochastic processes in network measurements

Markku Liinaharja (slides originally made by Esa Hyytiä),

Department of Communications and Networking
Aalto University, School of Electrical Engineering

Version 0.2, October 4, 2017

Contents

Stochastic process - definition

- Mapping from the sample space to a function
  - Each random outcome corresponds to a realization of a stochastic process that is a function of "time"
- Sequence of random variables
  - Continuous-time processes $\{X_t,\ t \geq 0\}$
  - Discrete-time processes $\{X_n,\ n = 1, 2, \ldots\}$
- Defined by the n:th order distributions (for all n), e.g.,
  - First-order: $\{f_{X_1}(x), f_{X_2}(x), \ldots\}$
  - Second-order: $\{f_{X_1,X_2}(x), f_{X_1,X_3}(x), \ldots\}$
- All the distributions are needed to include all possible dependencies

Stationary processes

- Stochastic process is (strictly) **stationary** if all of its distribution functions are invariant under time shifts
- Simpler and more useful conditions, e.g.,
  - Wide-sense/weakly stationary process: mean and autocovariance are invariant in time
- 1st order statistics, mean
  - Expectation at time t, $\mathrm{E}(X_t)$
- 2nd order statistics, autocovariance

$$R_{t,s} = \mathrm{E}((X_t - E[X_t])(X_s - E[X_s])) = \mathrm{Cov}[X_t, X_s].$$

- Wide-sense discrete-time stationary process, for all n, k:

$$\mathrm{E}(X_n) = \mathrm{E}(X_1), \quad \mathrm{Cov}[X_n, X_{n+k}] = \mathrm{Cov}[X_1, X_{k+1}].$$

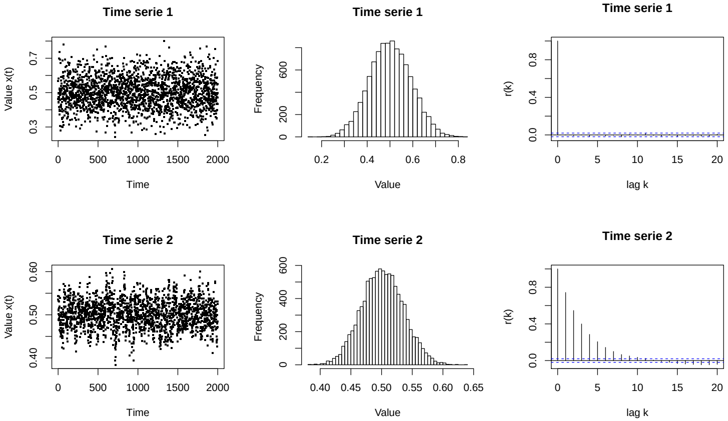
Autocorrelation

- If a process is at least wide-sense stationary, we can specify useful properties without regard to the particular time instant n
- Autocovariance of a wide-sense stationary process depends only on the lag k between random variables X_i and X_{i+k}
- Autocorrelation is normalized autocovariance

$$r(k) = \frac{\mathrm{Cov}[X_n, X_{n+k}]}{\mathrm{Cov}[X_n, X_n]}.$$

Example: autocorrelation (acf)

Time serie | **Histogram** | **acf**

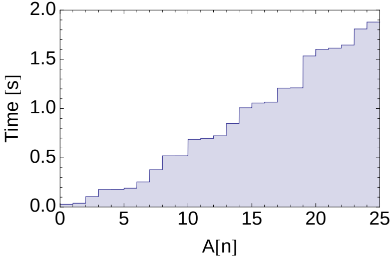


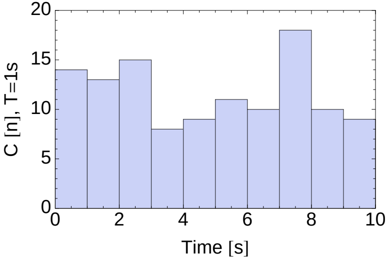
Samples in time serie 1 appear to be i.i.d. (no dependence)

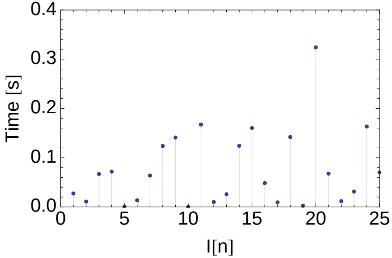
Some stochastic processes of interest

- **Arrival process**
  - $\{A_n,\ n = 0, 1, \ldots\}$, where A_n is the time instant of the nth arrival
  - Non-decreasing, non-stationary
- **Inter-arrival process**
  - $\{I_n,\ n = 1, 2, \ldots\}$, where $I_n = A_n - A_{n-1}$ is the length of the nth inter-arrival time
- **Time series of counts**
  - For a fixed time interval T
  - Time series of counts $\{C_n,\ n = 0, 1, \ldots\}$, where $C_n = \#\{A_m | nT < A_m < (n+1)T\}$
  - Arrivals can be packets, bits, bytes, ...
  - Most common form of reporting network traffic

Example: session start times





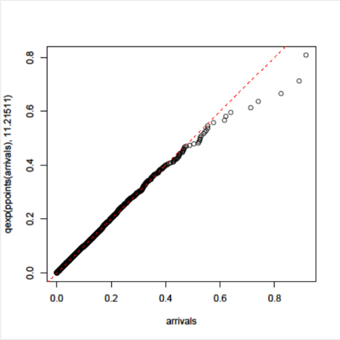


Poisson process

- Inter-arrival process of **Poisson process** is such that
  - I_n and I_k are **independent** for $n \neq k$
  - I_n obeys Exp(λ) distribution for some parameter λ
- Alternatively, but equivalently: Poisson process is such that for any $T > 0$
  - C_n and C_k are independent for all $n \neq k$
  - $C_n \sim$Poisson(λT)
- Poisson process is a widely used arrival process in performance modeling and analysis
  - "Infinitely" large population from which arrivals come independently
  - Often a valid assumption for e.g. human behavior in a large population
- Good for session arrivals (and phone calls ...)

Example

- Are the session start times from a Poisson process?
- Independent and exponentially distributed inter-arrivals?
- ACF, qqplots, ecdf, ...




```
> arrivals<-scan("testidata.txt");
Read 4360 items
> fitdistr(arrivals,"exponential");
   rate
 11.2151101
( 0.1698479)
> qqplot(arrivals, qexp(ppoints(arrivals), 11.21511));
abline(0,1, col = 2, lty = 2)
```

Poisson conversation . . .

Packets arrive according to a Poisson process with rate . . .

But in reality the arrival process is NOT POISSONIAN!!

Dependence structure

- As a single flow can contribute packet or byte arrivals to several consecutive measurement periods, the independence assumption of Poisson process breaks down, e.g., for packet count processes
- If ACF plot reveals that past values contribute to the present other models must be considered
  - Note again that stationary processes are only good for describing stable data ...
  - Network traffic is generally not stable at long time scales

Short- and Long-Range Dependence

1. **Short-range dependence** (SRD):

   The coupling between values at different times decreases rapidly as the lag k increases

   No tail: $r(k) = 0$ for $k > \theta$, or
   Exp-tail: $r(k) \sim \beta^{-k}$ for $k > \theta$, (for some $\theta > 0$ and $\beta > 1$)

2. **Long-range dependence** (LRD)

   The coupling between values at different times decreases slower than exponentially

   Power-law tail: $r(k) \sim k^{-\beta}$, for $k > \theta$ (for some $\theta>0$ and $0<\beta<1$)

SRD: $\sum_{k=1}^{\infty} r(k) < \infty$	LRD: $\sum_{k=1}^{\infty} r(k) = \infty$

Time series modeling

- Time series analysis provides suitable models for situations where autocorrelation shows significant dependence
- Examples from models
  - **Moving average (MA)**

$$Y_n = \mu + X_n + a_1 X_{n-1} + a_2 X_{n-2} + \ldots + a_k X_{n-k}$$

    where $\mathrm{E}(X_n) = 0$ and $\mathrm{E}(X_n^2) = \sigma^2$.
  - **Autoregressive (AR)**

$$Y_n = X_n + a_0 + a_1 Y_{n-1} + \ldots + a_k Y_{n-k}$$

    where $\mathrm{E}(X_n) = 0$, $\mathrm{E}(X_n^2) = \sigma^2$, and $\mathrm{E}(X_n X_m) = 0$ for $n \neq m$.
  - **EWMA, ARMA, ARIMA, DAR, ...**
- Well-developed field and especially widely used in signal processing, econometrics, etc.
- Cf., `http://www.statsoft.com/textbook/sttimser.html`

Contents

- Stochastic processes
- $\boxed{\text{Self-similarity}}$

Scaling

Long-range dependence is closely linked to a special property of network traffic; its **scaling behavior**

Rescaled view of traffic (e.g. time series of byte counts)

$$X_n^{(m)} \triangleq \sum_{i=nm}^{nm+m-1} X_i.$$

If the original data was observed on a time scale T, the rescaled process is observed on time scale Tm

Self-similarity

- A zero mean stochastic process $\{X_n\}$ is called **self-similar** with **Hurst parameter H** if, for all m, the aggregated process $\{X^{(m)}\}$ has the same distribution as $\{m^H X_n\}$, i.e.,

$$\{m^H X_n\} \stackrel{d}{=} \{X^{(m)}\}, \quad m > 0 \text{ and } 1/2 \leq H < 1.$$

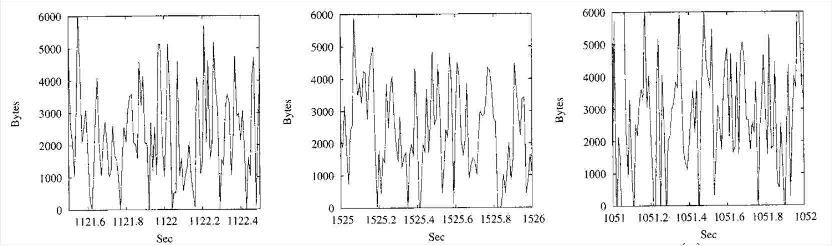
- Process can be self-similar only if it is LRD:
  - In fact, asymptotically autocorrelation $r(k) \sim k^{2H-2}$
  - Self-similarity means that $r^{(m)}(k) = r(k)$, with $k, m > 0$

  "Self-similar processes have the same autocorrelation function on all time scales"

- *(Nonself-similar: $r^{(m)}(k) \to 0$ as $m \to \infty$ for $k = 0, 1, 2, \ldots$)*
- Self-similarity describes how the variability of the process scales

Self-similarity in Network Traffic

Real traffic from Internet Traffic Archive | Synthetic traffic with $H = 0.83$ | Synthethic traffic with $H = 0.5$

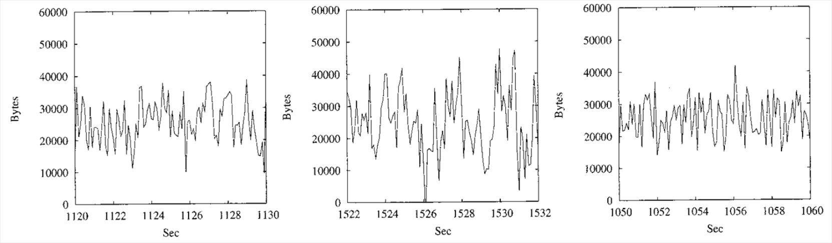


Timescale: 10 ms

Source: Crovella, Krishnamurthy: Internet Measurement: Infrastructure, Traffic & Applications, Wiley, 2006.

Self-similarity in Network Traffic

Real traffic from Internet Traffic Archive	Synthetic traffic with $H = 0.83$	Synthethic traffic with $H = 0.5$

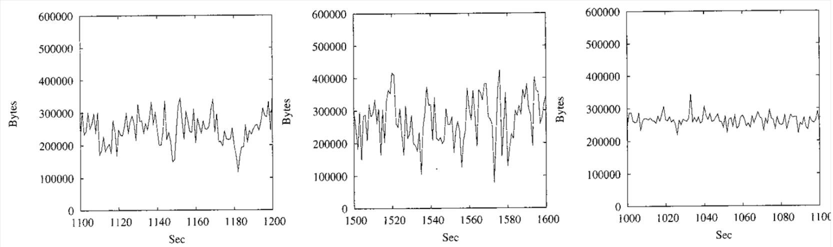


Timescale: 100 ms

Source: Crovella, Krishnamurthy: Internet Measurement: Infrastructure, Traffic & Applications, Wiley, 2006.

Self-similarity in Network Traffic

Real traffic from Internet Traffic Archive | Synthetic traffic with $H = 0.83$ | Synthethic traffic with $H = 0.5$



Timescale: 1 s

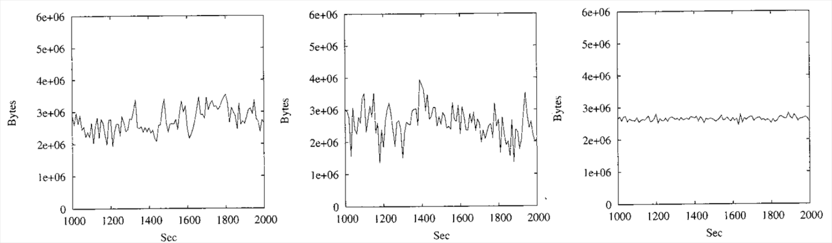
Source: Crovella, Krishnamurthy: Internet Measurement: Infrastructure, Traffic & Applications, Wiley, 2006.

Self-similarity in Network Traffic

Real traffic from Internet Traffic Archive

Synthetic traffic with $H = 0.83$

Synthethic traffic with $H = 0.5$



Timescale: 10 s

Source: Crovella, Krishnamurthy: Internet Measurement: Infrastructure, Traffic & Applications, Wiley, 2006.

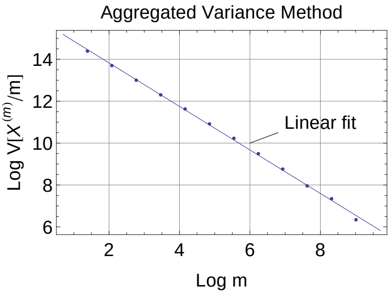
Measuring self-similarity

- Self-similarity results from variety of factors, but the main cause is the long-range dependency caused by heavytailed file/flow-size distributions
- Simplest way of estimating the Hurst parameter H is called the **aggregated variance** method:
  - Let X_n be a series of counts
  - Plot the variance of $m^{-1}X^{(m)}$ against m on a log-log scale
  - If the original data is well modeled with a self-similar process, then the variance of $m^{-1}X^{(m)}$ will follow m^{2H-2}

    Plot should show a straight line with slope $-\beta$ greater than -1
  - $H = 1 - \beta/2$
- Other, more efficient, methods exist but are beyond the scope of our course

Example (simple)

- Let X_n be i.i.d. random variables, $X_n \sim \text{Pareto}(100, 1.5)$
- I.e., infinite variance, with a finite mean $\text{E}(X) = 300$
- Compute aggregates $X^{(m)}/m$ for $m = 2^1, 2^2, 2^3, \ldots$ and their sample variance $s^2_{(m)}$
- Mean remains the same, $\text{E}(X) = 300$
- Plot $\log m$ against $\log s^2_{(m)}$



Slope is -1, i.e., $H = 0.5$ and process is not self-similar(!)

Why self-similarity?

- Self-similarity is a parsimonious model to describe the variability of traffic on different time scales
- Self-similarity has fundamental effects on queuing behavior of traffic and hence it is interesting in performance analysis
- Recognition of the phenomenon is seen as a fundamental step forward in the way analysts think about traffic
- However, it is easy to mix with non-stationarity
  - Hurst parameter estimates from a non-stable data will be misleading

Literature

- Mark Crovella, Balachander Krishnamurthy, **Internet Measurement: Infrastructure, Traffic & Applications**, John Wiley and Sons Ltd, 2006.