Markovian Arrival Processes for Time Series Modeling

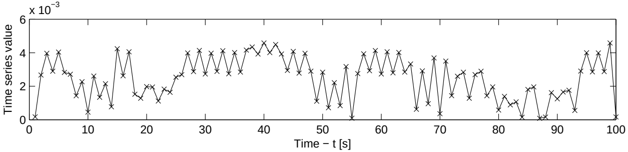
Giuliano Casale
casale@cs.wm.edu

Department of Computer Science
College of William and Mary

May 13th, 2008

Time Series Fitting Models

Motivating problem : model a sequence of data points as a stochastic process X_t with known mathematical structure



Model Requirements :

1. Capture the distribution (y-axis value)
2. Capture the temporal dependence (x-axis placement)
3. Analytical tractability
4. Integration in larger models
5. Compactness
6. Scalability of fitting algorithms

Some Popular Time Series Models...

Markovian models

1. Phase-type (PH-type) renewal processes (no order)
2. Markov-modulated processes (MMPPs, MAPs)

Gaussian models

1. Brownian motion (RBM, FBM), stochastic calculus
2. Autoregressive models (AR, ARMA, ARIMA, FARIMA,. . .)

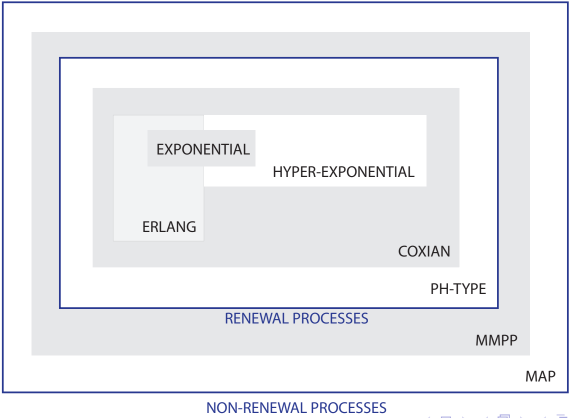
Signal-processing methods

1. Fourier transforms
2. Wavelets

Taxonomy

Hierarchy of processes with increasing fitting capabilities :

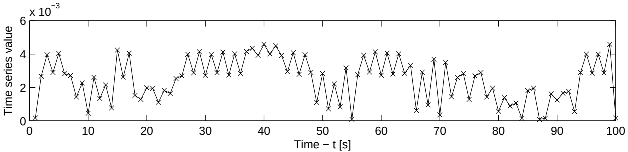
- Renewal models : capture distribution, ignore sample order
- Nonrenewal models : capture both distribution and ordering

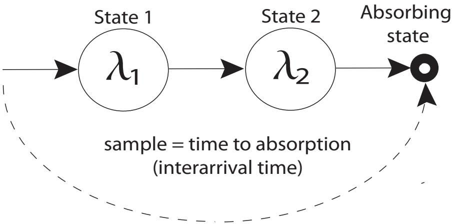


Outline

1. Markovian Fitting Models
2. Renewal Processes (Distribution Fitting)
3. Markovian Arrival Processes (MAPs)
4. Kronecker Product Composition (KPC)

Markovian Fitting Models





Markovian Fitting Models

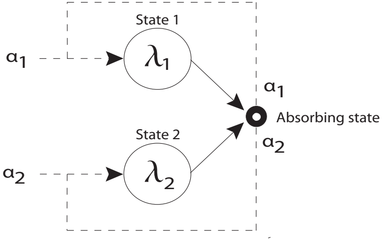
Basic idea :

- we consider a Markov chain with an absorbing state
- initial state is state k with probability α_k
- data point $\equiv$ time X_t between initialization and absorption
- time series $\equiv$ sequence of initialization$\rightarrow$absorption cycles

Fitting :

- determine state space and jump rates such that X_t has same distribution and temporal dependence of the time series
- the Markov chain is then a model of the time series

Example : Hyper-exponential renewal process

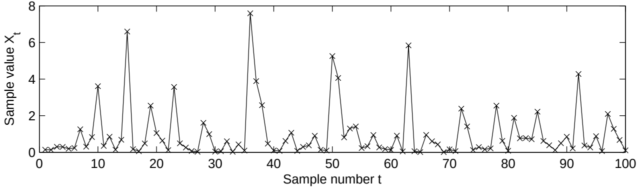


Model :

$$Q = \begin{bmatrix} -\lambda_1 & 0 & \lambda_1 \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 \end{bmatrix}, \vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3 \equiv 0), \alpha_1 + \alpha_2 = 1$$

- PDF : $Pr[X_t = x] = \alpha_1 \lambda_1 e^{-\lambda_1 x} + \alpha_2 \lambda_2 e^{-\lambda_2 x}$
- Moments : $E[X_t^k] = k![\alpha_1(-\lambda_1)^{-k} + \alpha_2(-\lambda_2)^{-k}]$
- Renewal model : $Pr[X_t | X_{t-k}] = Pr[X_t], \forall k$

Fitting a Hyper-exponential renewal process



Empirical Moments : $E[S] = 0.981,\ E[S^2] = 3.059,\ E[S^3] = 14.83$
Model fitting by moment matching $(E[X_t^k] \equiv E[S^k])$:

$$\begin{cases} [\alpha_1(-\lambda_1)^{-1} + \alpha_2(-\lambda_2)^{-1}] = E[S] \\ 2[\alpha_1(-\lambda_1)^{-2} + \alpha_2(-\lambda_2)^{-2}] = E[S^2] \\ 6[\alpha_1(-\lambda_1)^{-3} + \alpha_2(-\lambda_2)^{-3}] = E[S^3] \\ \alpha_1 + \alpha_2 = 1 \end{cases}$$

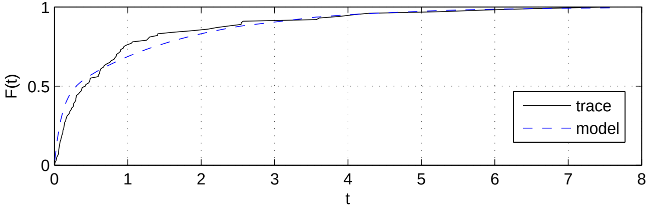
Solution $(\lambda_1 = 0.0932, \lambda_2 = 1.6197, \alpha_1 = 0.4184, \alpha_2 = 0.5816)$

Goodness-of-fit

Fitted model :

$$Q = \begin{bmatrix} -0.0932 & 0 & 0.0932 \\ 0 & -1.6197 & 1.6197 \\ 0 & 0 & 0 \end{bmatrix}, \vec{\alpha} = (0.4184, 0.5816, 0)$$

Empirical vs model cumulative distribution function :



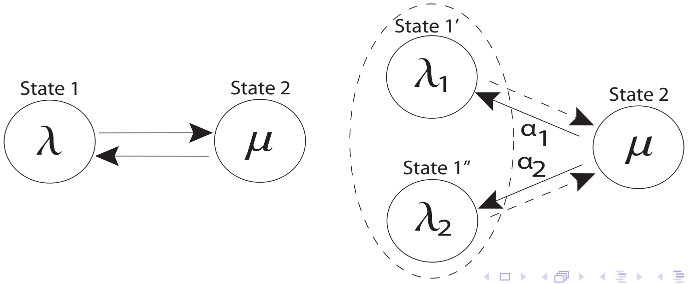
Integration in existing Markov models

Motivation : hyper-exponential times X_t for a transition $A \rightarrow B$

Integration steps :

- model X_t as a small Markov chain with states s_k, $1 \leq k \leq K$
- augment state A to track the value of s_k
- use B in place of the absorbing state of X_t
- modify the transitions $B \rightarrow A$ to initialize X_t according to α

Integration Example :



Markovian Fitting Models : Pros and Cons

Advantages :

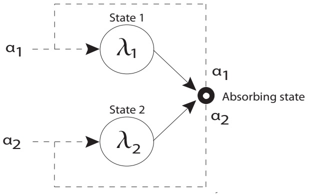
- Markov models are analytically tractable
- Extensive literature available (also textbooks)
- Easy to integrate in larger Markov models
- Important distributions are simple to model (e.g., Erlang, hyper-exponential)
- Support for temporal dependence in a Markovian setting

Drawbacks :

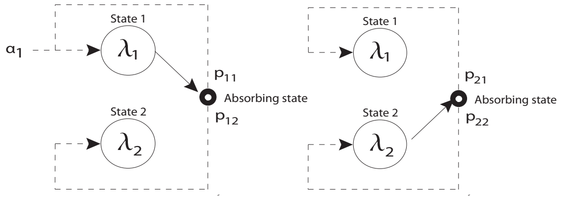
- Lack of compactness : for general distribution and state space size can be large (16-32 states)
- Fitting algorithms are still investigated

Markovian Arrival Processes (MAPs)

PH-type Renewal Process :



Markovian Arrival Process :



Markovian Arrival Processes (MAPs)

MAPs subdue all fitting models based on Markov chains

Features :

- The state space is a general Markov chain ($\Rightarrow$ MAP samples are PH-type distributed)
- MAPs make each sample dependent on the actions taken to generate the last sample ($\Rightarrow$ temporal dependence). In a MAP, after absorption from state k, we restart from state i with fixed probability $p_{k,i}$. Thus, the initialization vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_K)$ is replaced by a transition matrix

$$P = [p_{k,i}], \quad 1 \leq k \leq K, 1 \leq i \leq K,$$

which depends on past history (i.e., state k).

MAP representation

$\mathrm{MAP}(K) \equiv$ MAP with K states (excluding absorbing state)

$$Q = \begin{bmatrix} -\lambda_1 & 0 & \lambda_1 \\ 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 \end{bmatrix}, P = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix},$$

Equivalent (D_0, D_1) description :

- D_0 represents transitions not directed to the absorbing state
- $D_1 = -D_0 P$ includes absorption + initialization information

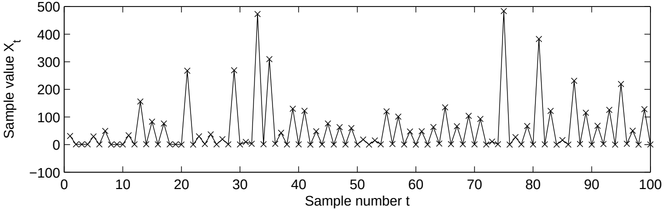
$$D_0 = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}, D_1 = \begin{bmatrix} p_{1,1}\lambda_1 & p_{1,2}\lambda_1 \\ p_{2,1}\lambda_2 & p_{2,2}\lambda_2 \end{bmatrix},$$

"Alternating" Hyper-exponential MAP

Simplest example of temporal dependent MAP :

$$D_0 = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}, D_1 = \begin{bmatrix} 0 & \lambda_1 \\ \lambda_2 & 0 \end{bmatrix}$$

Sampling alternates between an exponential with mean $\lambda_1 = 1$ and an exponential with mean $\lambda_2 = 100$.



Consecutive samples (lag one) are *negatively* correlated in magnitude, samples spaced by two lags are *positively* correlated.

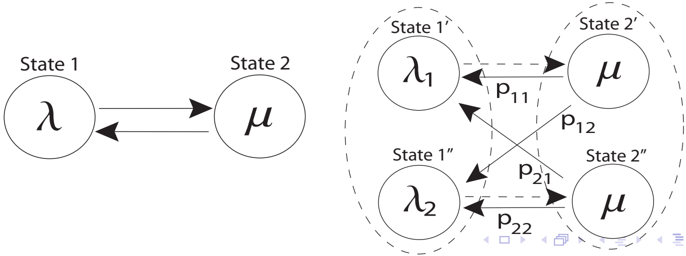
Integration of MAPs in existing Markov models

Motivation : MAP times X_t for a transition $A \rightarrow B$

Integration steps :

- model X_t as a small Markov chain with states s_k, $1 \leq k \leq K$
- augment state A to track the value of s_k
- use B in place of the absorbing state of X_t
- augment state B to track the last visited state
- modify the transitions $B \rightarrow A$ to initialize X_t according to P

Integration Example :



MAP Distribution

Define $\vec{\pi}_e = \vec{\pi}_e P = \vec{\pi}_e(-D_0)^{-1}D_1$, $\vec{1} = (1, 1, \ldots, 1)$.
Moments :

$$E[X^k] = k!\vec{\pi}_e(-D_0)^{-k}\vec{1}$$

Cumulative distribution function (CDF) :

$$F(x) = 1 - \vec{\pi}_e e^{D_0 x}\vec{1}$$

Probability density function (PDF) :

$$f(x) = 1 - \vec{\pi}_e e^{D_0 x}(-D_0)\vec{1}$$

Fitting methods for MAP distribution (i.e., assign D_0 and $\vec{\pi}_e$) :

- Inversion of closed-form formulas (e.g., moment matching)
- EM-algorithms

MAP Temporal Dependence

Temporal dependence : the distribution of X_t depends on the past sampled values $X_{t-1}, X_{t-2}, \ldots, X_{t-k}$. In MAPs we restrict our attention to finite k.

Frequently used descriptions of temporal dependence :

- Joint probabilities : $\Pr[X_t, X_{t-1}, \ldots, X_{t-k}]$
- Joint moments : $E[X_t^{a_0} X_{t-1}^{a_1} \cdots X_{t-k}^{a_k}]$
- Normalized joint moments : e.g., autocorrelation coefficients

$$\rho_k = \frac{E[X_t X_{t-k}] - E[X]^2}{E[X^2] - E[X]^2}$$

MAP Temporal Dependence

Temporal dependence can be described in MAPs by closed-form formulas depending on D_0 and D_1.
Joint probabilities :

$$\Pr[X_t = x_0, \ldots, X_{t-k} = x_k] = \vec{\pi}_e(\prod_{i=0}^{k} e^{D0 x_i} D_1)\vec{1}$$

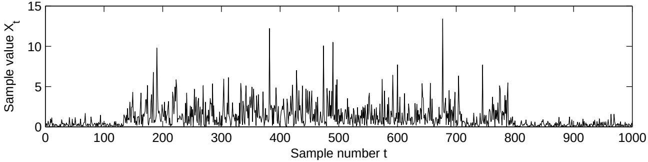
Joint moments :

$$E[X_t^{a_0} X_{t-1}^{a_1} \cdots X_{t-k}^{a_k}] = \vec{\pi}_e(\prod_{i=0}^{k} a_k!(-D_0)^{-a_k-1} D_1)\vec{1}$$

Autocorrelation coefficients :

$$\rho_k = \frac{\vec{\pi}_e[(-D_0)^{-1} D_1]^k (-D_0)^{-1}\vec{1} - E[X_t]}{2\vec{\pi}_e(-D_0)^{-1}\vec{1} - E[X_t]}$$

Fitting a time series by a Hyperexponential MAP(2)



$$D_0 = \begin{bmatrix} -\lambda_1 & 0 \\ 0 & -\lambda_2 \end{bmatrix}, D_1 = \begin{bmatrix} p_{1,1}\lambda_1 & p_{1,2}\lambda_1 \\ p_{2,1}\lambda_2 & p_{2,2}\lambda_2 \end{bmatrix}$$

We determine the MAP(2) parameters by matching $E[S], E[S^2], E[S^3]$, imposing $p_{1,1} + p_{1,2} = 1$, $p_{2,1} + p_{2,2} = 1$, and matching the autocorrelation function.

When there are only two states, this is given by :

$$\rho_k = \frac{1}{2}(1 - E[X_t]^2(E[X_t^2] - E[X_t]^2)^{-1})\gamma_2^k, \quad \gamma_2 = 1 - p_{2,1} - p_{1,2},$$

thus we need only to impose γ_2.

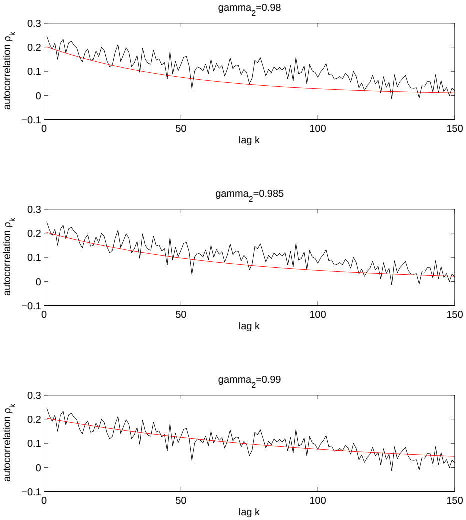
Fitting a time series by a Hyperexponential MAP(2)

Time Series : $E[S] = 0.846,\ E[S^2] = 2.445,\ E[S^3] = 15.03,$
$\rho_1 = 0.2898$
Model fitting by moment and autocorrelation matching :

$$\begin{cases} [\frac{p_{2,1}}{p_{2,1}+p_{1,2}}(-\lambda_1)^{-1} + \frac{p_{1,2}}{p_{2,1}+p_{1,2}}(-\lambda_2)^{-1}] = E[S] \\ 2[\frac{p_{2,1}}{p_{2,1}+p_{1,2}}(-\lambda_1)^{-2} + \frac{p_{1,2}}{p_{2,1}+p_{1,2}}(-\lambda_2)^{-2}] = E[S^2] \\ 6[\frac{p_{2,1}}{p_{2,1}+p_{1,2}}(-\lambda_1)^{-3} + \frac{p_{1,2}}{p_{2,1}+p_{1,2}}(-\lambda_2)^{-3}] = E[S^3] \\ \frac{1}{2}(1 - E[X_t]^2(E[X_t^2] - E[X_t]^2)^{-1})(p_{2,1} + p_{2,2}) = \rho_1 \end{cases}$$

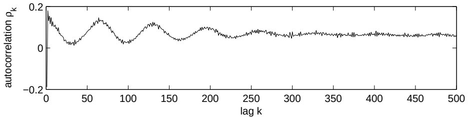
Solution $(\lambda_1 = 0.4183, \lambda_2 = 1.9273, p_{2,1} = 0.0083, p_{1,2} = 0.0017)$

Fitting a time series by a Hyperexponential MAP(2)



General MAP(K) Fitting

Real time series can show complex distribution or correlations that cannot be fitted accurately by simple two state models.



No MAP(2) can fit the above autocorrelation, because in MAP(2)

$$\rho_k \propto \gamma_2^k, \gamma_2 \in \Re$$

If $K > 2$ we can fit complex temporal dependence patterns, but it becomes increasingly difficult to do moment matching because of the non-linearity of the systems and the large number of parameters involved. E.g., a MAP(4) requires assigning 28 parameters to set D_0 and D_1.

Kronecker Product Composition (KPC)

$MAP^a = \{D_0^a, D_1^a\}$, $MAP^b = \{D_0^b, D_1^b\}$, their KPC is the MAP

$$MAP^a \otimes MAP^b = \{D_0, D_1\} = \{-D_0^a \otimes D_0^b, D_1^a \otimes D_1^b\},$$

where $\otimes$ denotes the Kronecker product operator and D_0^a must be diagonal. For example, if the original processes have D_0 matrices

$$D_0^a = \begin{bmatrix} -a_{1,1} & 0 \\ 0 & -a_{2,2} \end{bmatrix}, D_0^b = \begin{bmatrix} -b_{1,1} & b_{1,2} \\ b_{2,1} & -b_{2,2} \end{bmatrix},$$

then the composition yields $D_0 = -D_0^a \otimes D_0^b$, i.e.,

$$D_0 = \begin{bmatrix} -a_{1,1}b_{1,1} & 0 & a_{1,1}b_{1,2} & 0 \\ 0 & -a_{2,2}b_{1,1} & 0 & a_{2,2}b_{1,2} \\ a_{1,1}b_{2,1} & 0 & -a_{1,1}b_{2,2} & 0 \\ 0 & a_{2,2}b_{2,1} & 0 & -a_{2,2}b_{2,2} \end{bmatrix}$$

KPC Properties

KPC is a divide and conquer approach to MAP(K) fitting. That is, the composed process has moments, correlations, joint moments, and autocorrelations that are in simple relation with those of MAP^a and MAP^b, e.g.,

$$E[X_t^k] = E_a[X_t^k] E_b[X_t^k]/k!, \quad (1)$$

$$E[X_t X_{t+k}] = E_a[X_t X_{t+k}] E_b[X_t X_{t+k}], \quad (2)$$

$$E[X_t X_{t+k} X_{t+k+j}] = E_a[X_t X_{t+k}] E_b[X_t X_{t+k} X_{t+k+j}] \quad (3)$$

KPC fitting problem : impose the moments and correlations of *small MAPs* such that their KPC process matches the moments and correlations of the trace.

KPC Toolbox

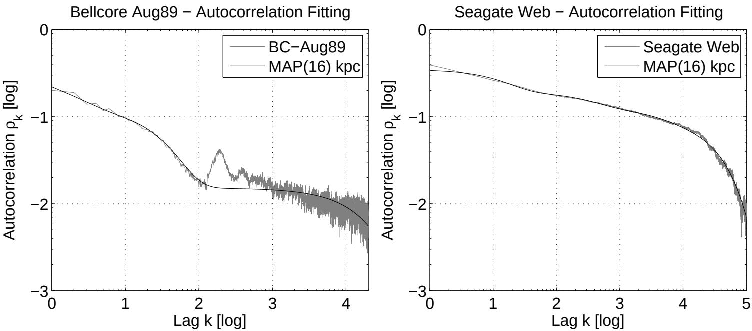
Input : the number J of MAP(2)s to be composed by KPC and the set of moments/correlations to be matched.

Fitting - Phase 1 : the non-linear optimization seeks for the set of autocorrelations of the small MAP(2)s that composed would minimize $|\rho_k^{KPC} - \rho_k|$.

Fitting - Phase 2 : the actual (D_0, D_1) representation of each MAP(2)s is determined. All residual degrees of freedom not assigned by Phase 1 are used to best match moments and higher-order correlations.

KPC Fitting Example

Fitting of real traces with 10^6 samples.



Good matching of ρ_k over many lags ($1 \leq k \leq 100,000$).
Distribution moments almost identical to the trace up to $E[X_t^5]$.

Conclusion

- Markovian fitting models can model accurately both distribution and order of samples
- Integration in larger Markov model is main motivation for their popularity
- MAP(2) models can be fitted accurately with closed-form formulas
- MAP(K) models currently investigated, KPC is the best-available fitting technique

Bibliography

- G. Casale, E. Z. Zhang, E. Smirni : Interarrival Times Characterization and Fitting for Markovian Traffic Analysis. William and Mary Tech. Rep., CS Dept, WM-CS-2008-02. http ://www.wm.edu/computerscience/techreport/2008/WM-CS-2008-02.pdf
- KPC Toolbox. `http://www.cs.wm.edu/MAPQN/kpctoolbox.html`
- G. Horváth, M. Telek, P. Buchholz : A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag correlation. QEST 2005 : 124-133
- M. Telek, G. Horváth : A minimal representation of Markov arrival processes and a moments matching method. Perform. Eval. 64(9-12) : 1153-1168 (2007)