Power Spectra from Symbol Sequences and ε-Machine Spectral Reconstruction Theory

Dowman P Varn

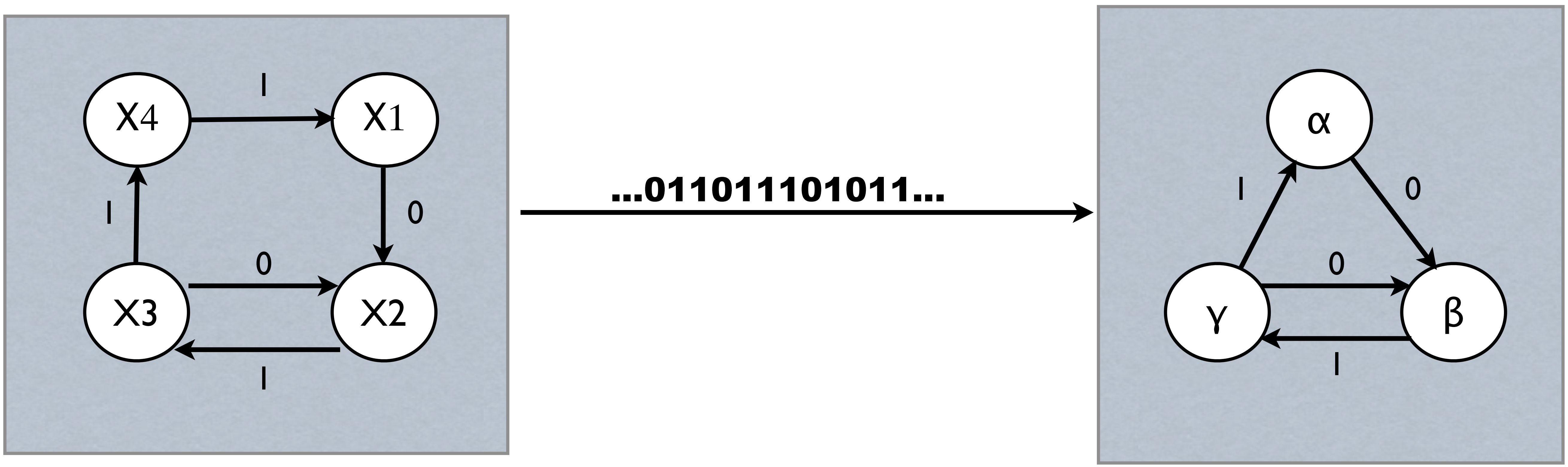
dpv@complexmatter.org

21 May 2013

Reading for this lecture:

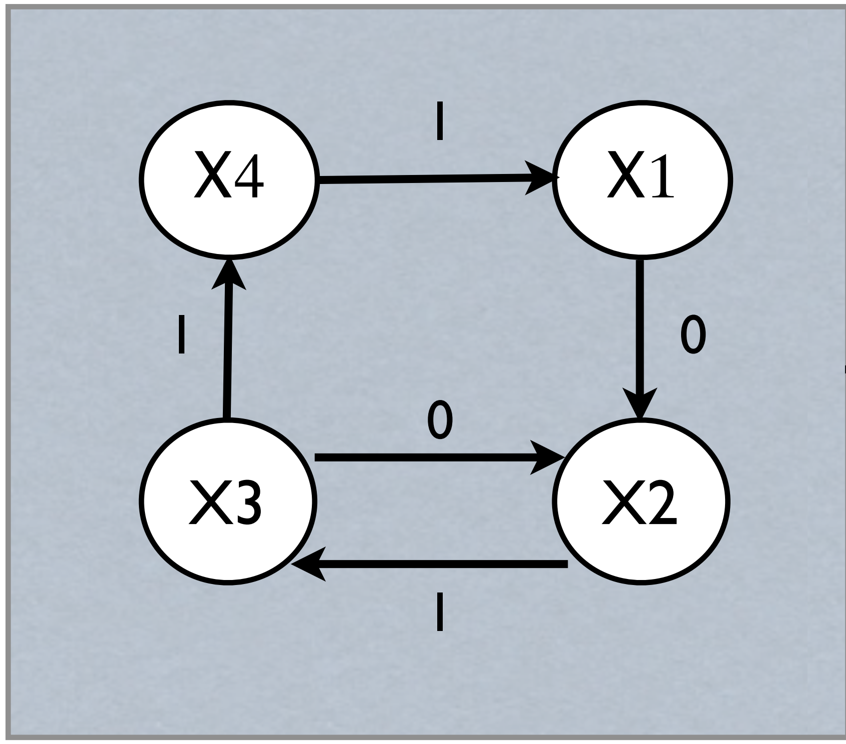
BTFM1 and BTFM2 articles in *CMR* and *Lecture Notes*

The Learning Channel I



System **Instrument** **Process** **Model**

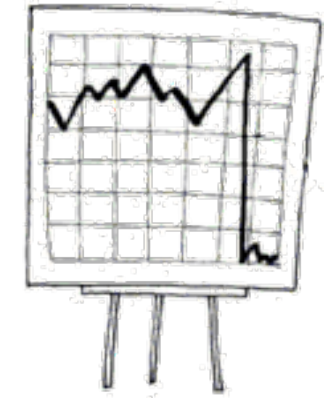
The Learning Channel II

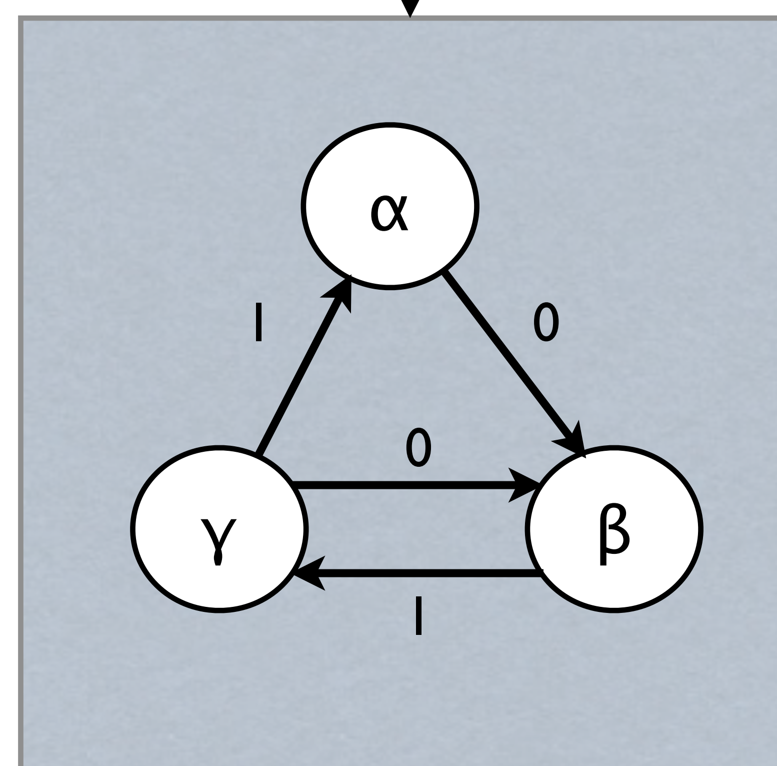


System

Signal Processing:
Power Spectrum

Instrument





Model

While the dynamics were hidden before and we could only observe the output sequences of 0s and 1s, now the signal is processed so we can't even observe that. We can can only see the Power Spectrum of the sequence.

Power Spectra of Discrete Series*

$$\mathsf{S}_N = s_0, s_1, \ldots, s_n, \ldots, s_{N-1} \qquad s_n \in \{0, 1\}$$

Define the *Discrete Fourier Transform* as:

$$\mathcal{F}(\mathsf{S}_N) = S(f) = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} e^{-2\pi i m f} s_m$$

The ***Power Spectrum*** is defined as:

$$\mathcal{P}(f) = |S(f)|^2$$

For purposes of computation, let's assign the numerical value of '1' to symbol '1', and the numerical value '-1' to symbol '0'. This choice is not unique.

*There is a much more elegant method of finding correlation functions and power spectra than demonstrated here. PM Riechers & JP Crutchfield have recently shown that the z-transformation can do much of this analytically, or if done numerically, much more efficiently.

The Correlation Function

Let's substitute the expression for the Fourier Transform into the for Power Spectrum:

$$\mathcal{P}(f) = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{m'=0}^{N-1} e^{-2\pi i f (m-m')} s_m s_{m'}$$

Now let *n* = *m* - *m'*

$$\mathcal{P}(f) = 1 + \frac{2}{N} \sum_{n=1}^{N-1} \sum_{m'=0}^{N-n-1} \cos(2\pi n f) s_{m'} s_{m'+n}$$

Define the *two-point correlation function* as

$$\mathcal{C}(n) \equiv \langle s_{m'} s_{m'+n} \rangle = \frac{1}{N-n} \sum_{m'=0}^{N-n-1} s_{m'} s_{m'+n}$$

$$\mathcal{P}(f) = 1 + \frac{2}{N} \sum_{n=1}^{N-1} (N-n)\mathcal{C}(n)\cos(2\pi n f)$$

This expression relating the power spectrum and the correlation function suggests that the latter can be found from Fourier analysis of the former, i.e.,

$$\mathcal{C}(n) = \int_0^1 \mathcal{P}(f)\cos(2\pi n f)\, df$$

This is a rather general result.

It is more convenient to work with the correlation functions in a slightly different form:

$$q(n) = \frac{1}{2}[\mathcal{C}(n) + 1]$$

q(n) is the probability that two symbols at distance *n* are identical.

A Classification Scheme for Power Spectra

Power Spectra can classified according to their scaling behavior with the length of the sequence.

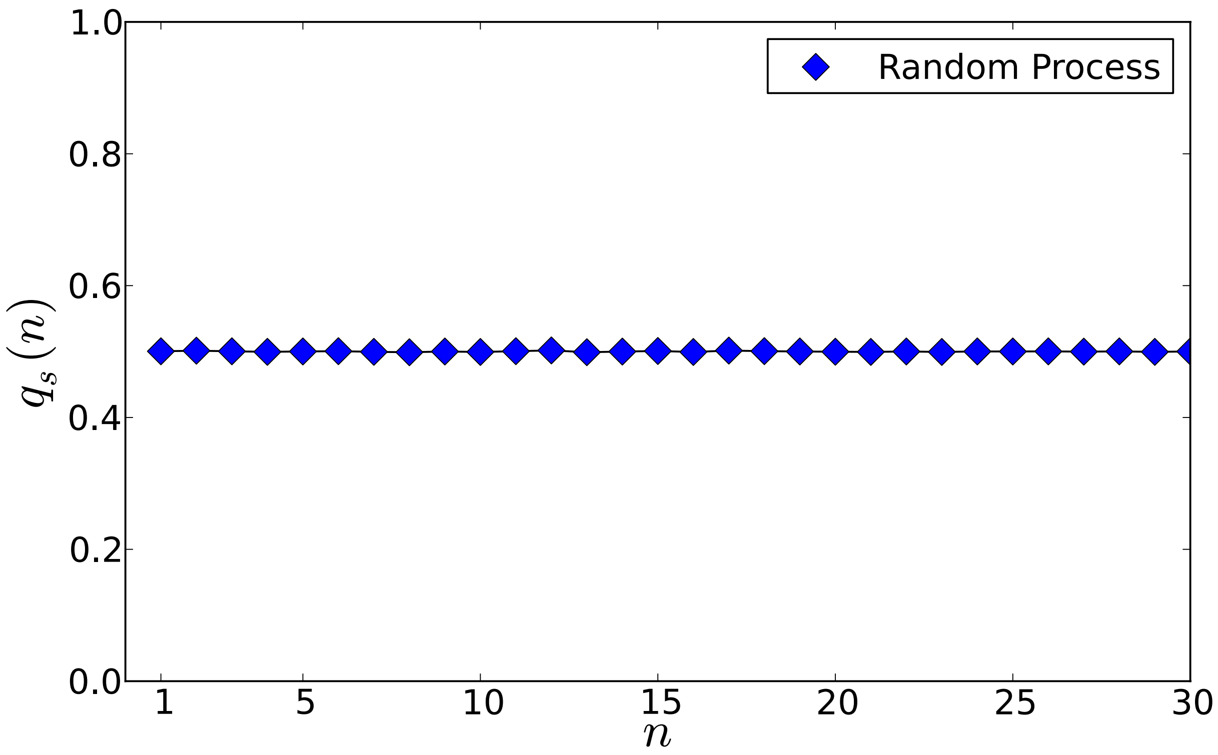
i. Pure Point $\mathcal{P}(f) \sim N^2$

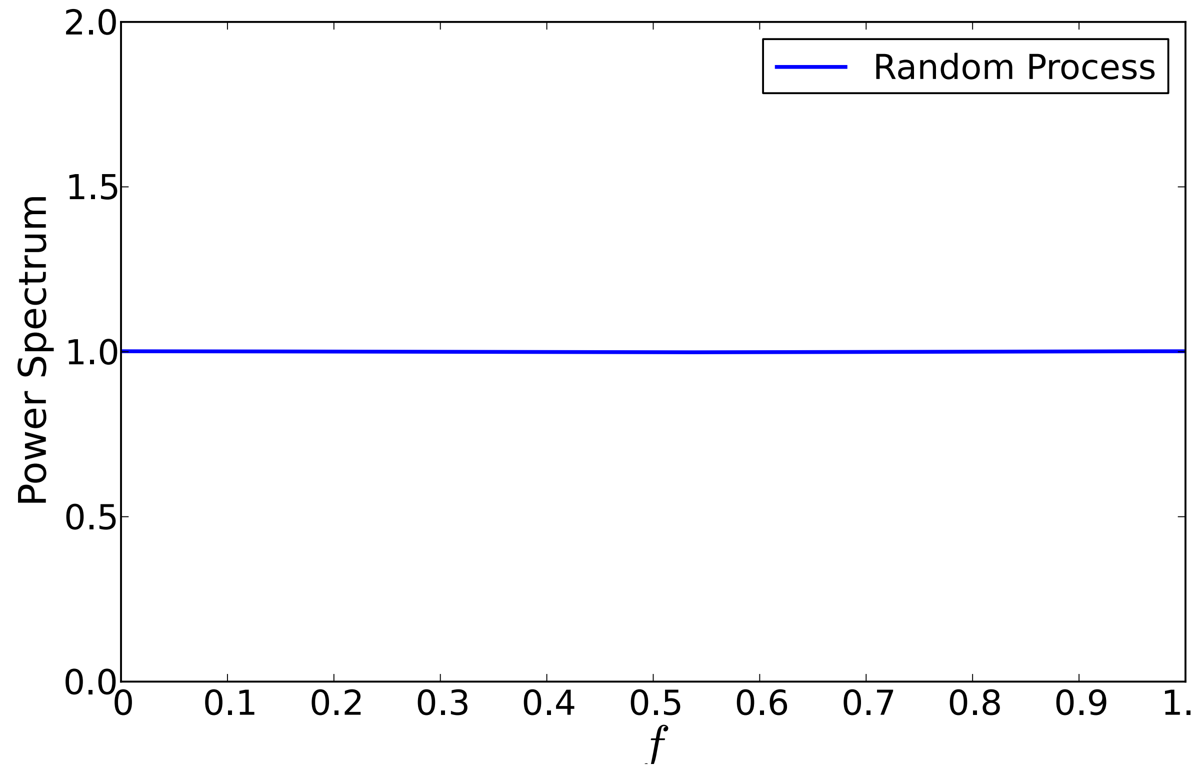
ii. Continuous $\mathcal{P}(f) \sim N$

iii. Singular Continuous $\mathcal{P}(f) \sim N^\gamma \quad 1 < \gamma < 2$

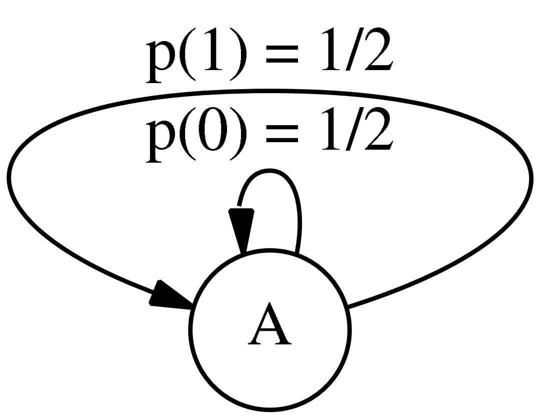
This scheme is rather basic, and certainly doesn't exhaust all the possibilities.
For instance, the power spectrum could scale like $N \log N$.

Examples of Power Spectra: Unbiased Coin Toss

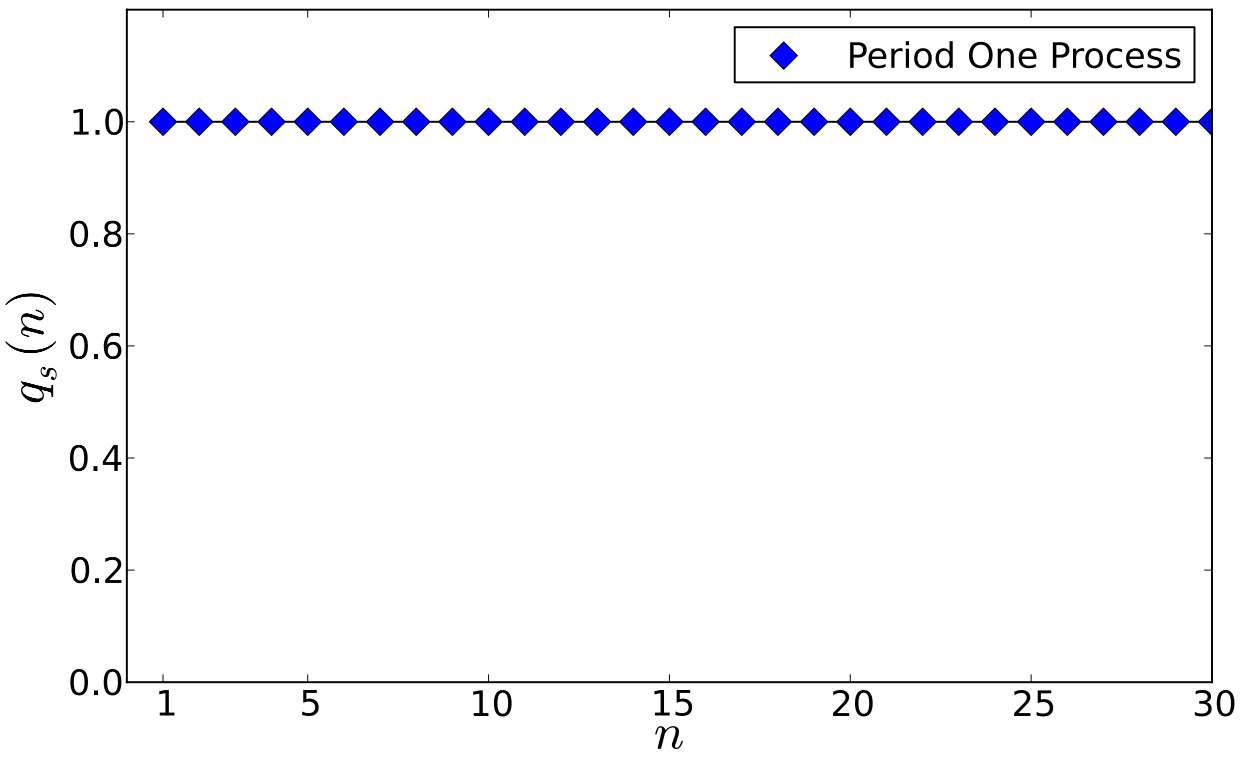


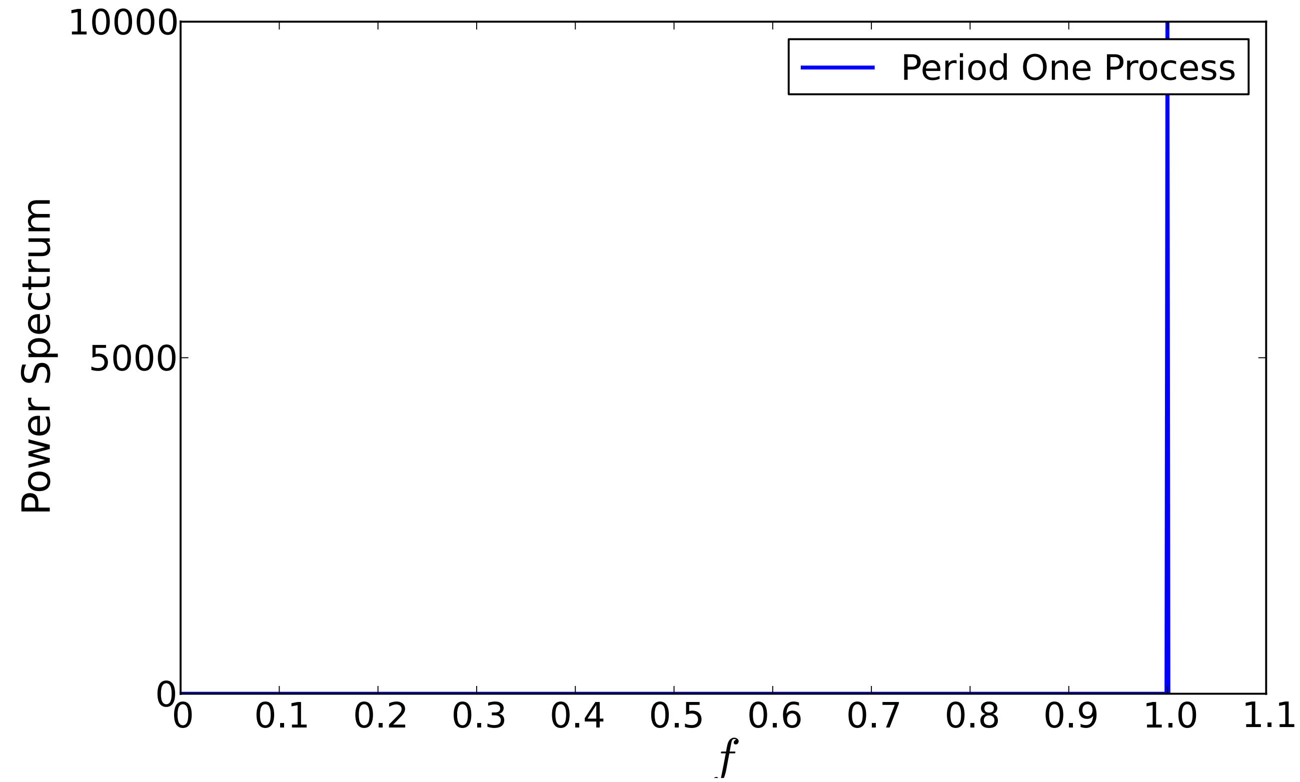


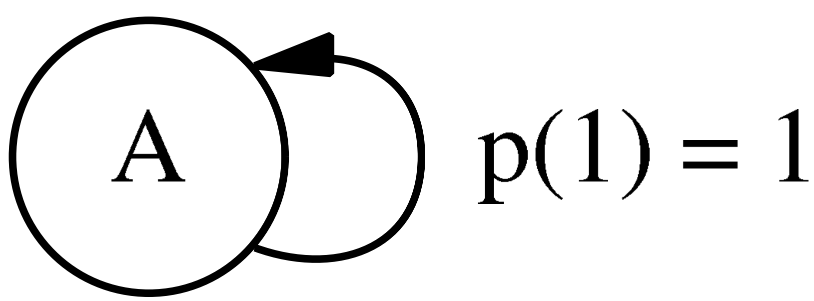
For a completely random sequence, both the correlation function and the power spectrum are featureless.



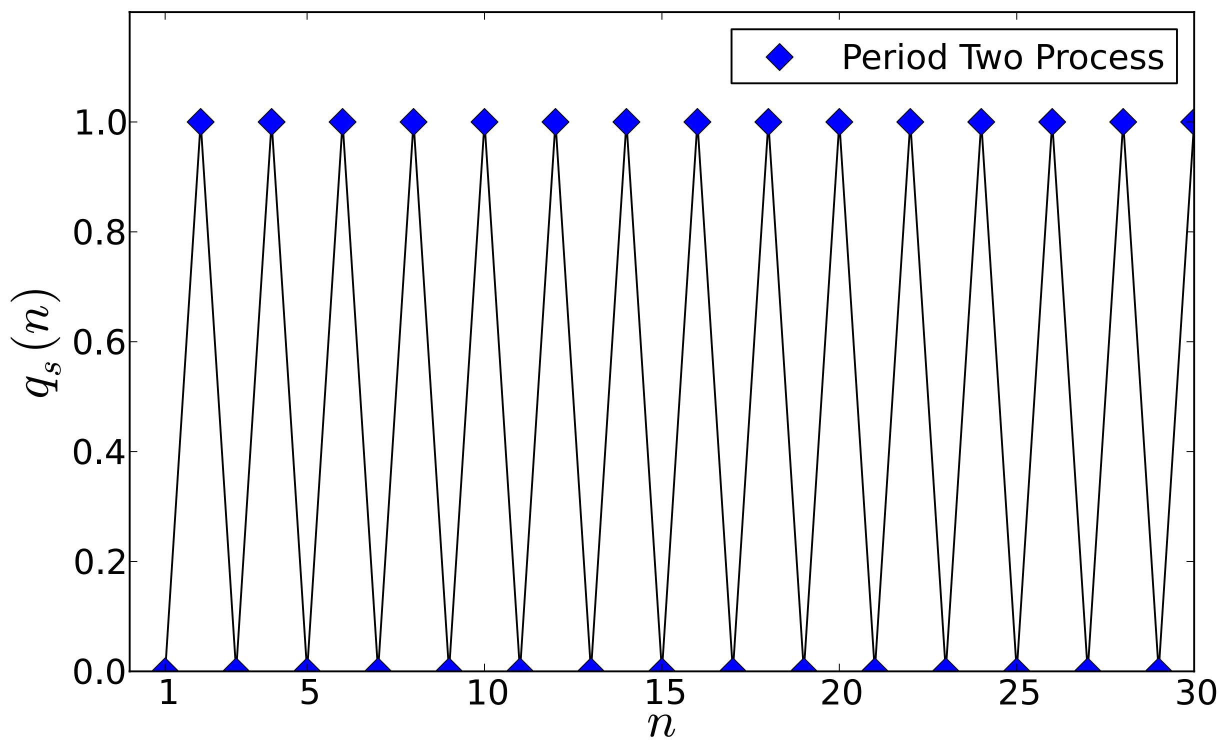
Examples of Power Spectra: Period 1

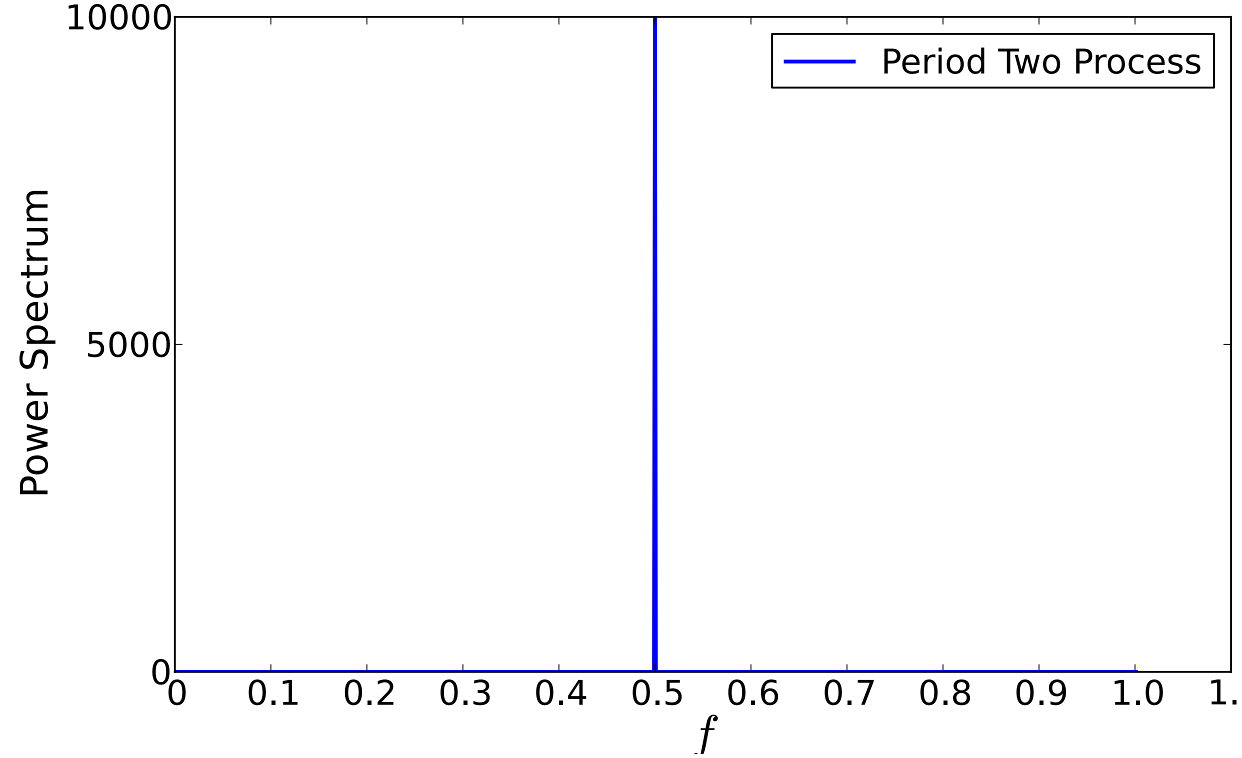


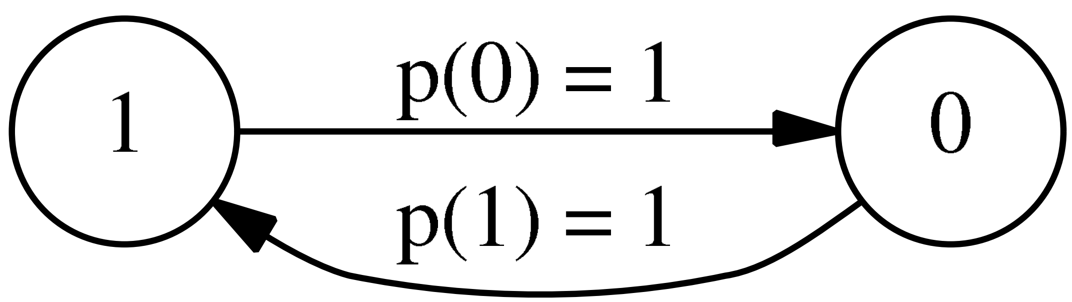




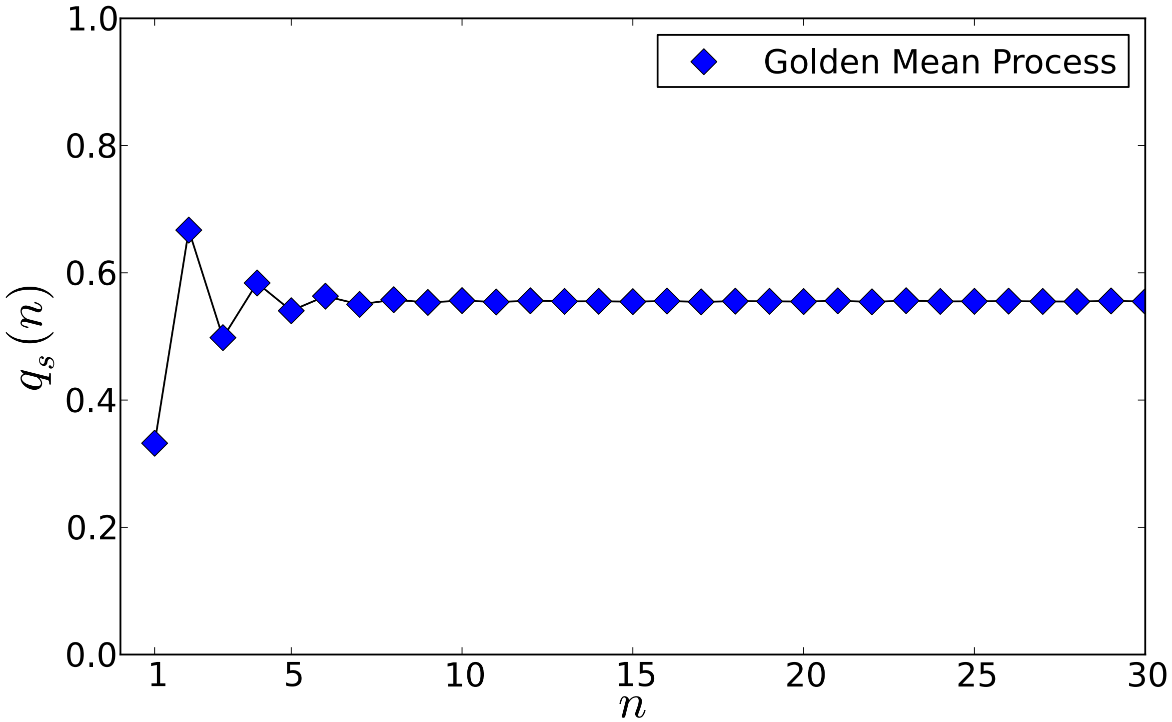
Examples of Power Spectra: Period 2

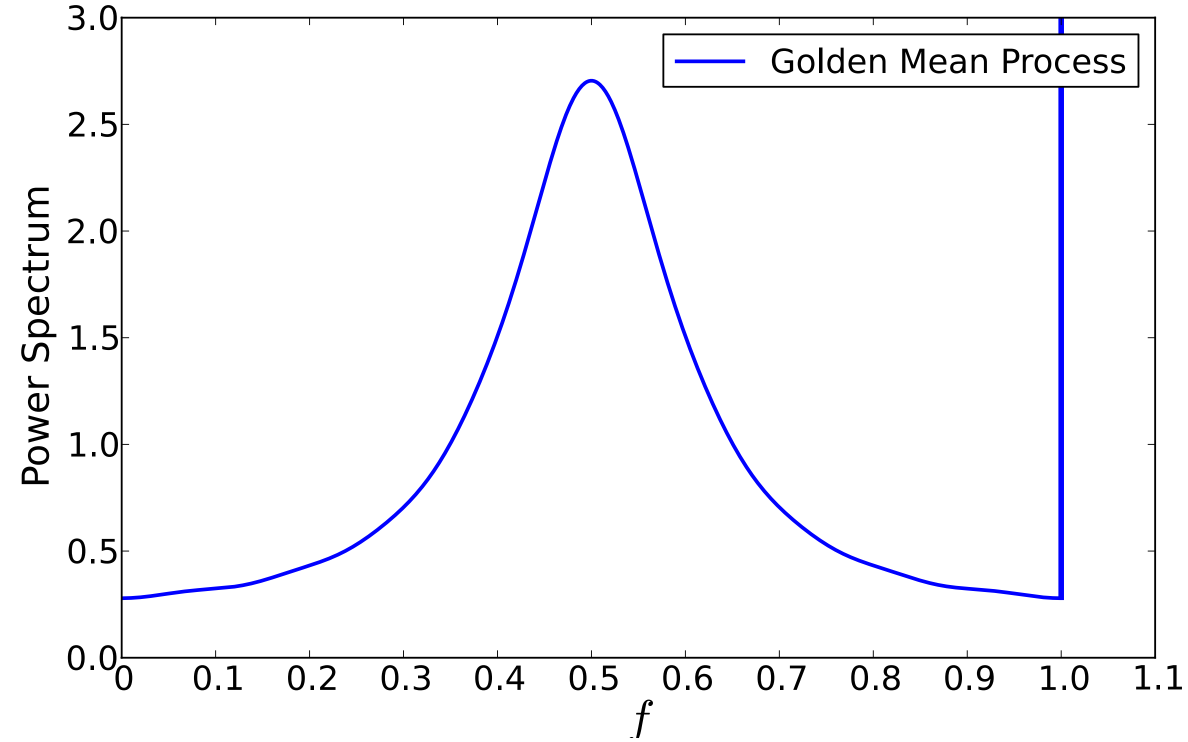


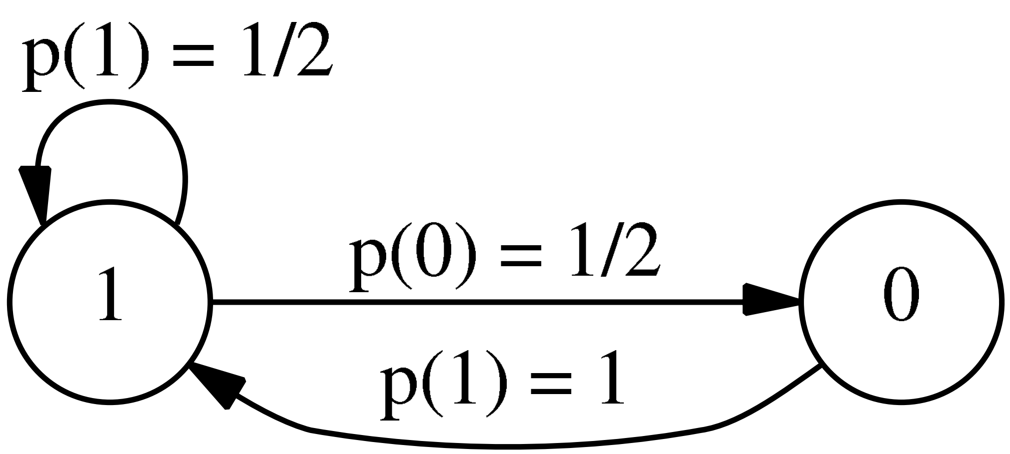




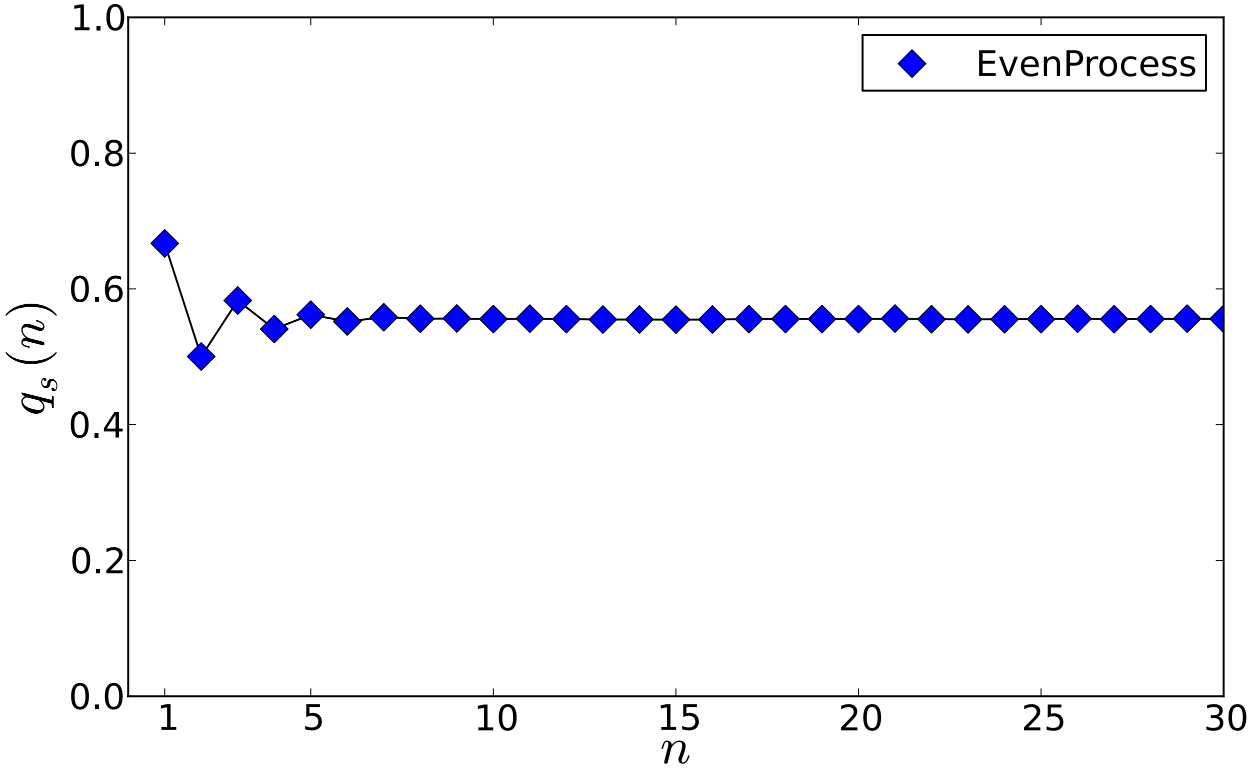
Examples of Power Spectra: Golden Mean

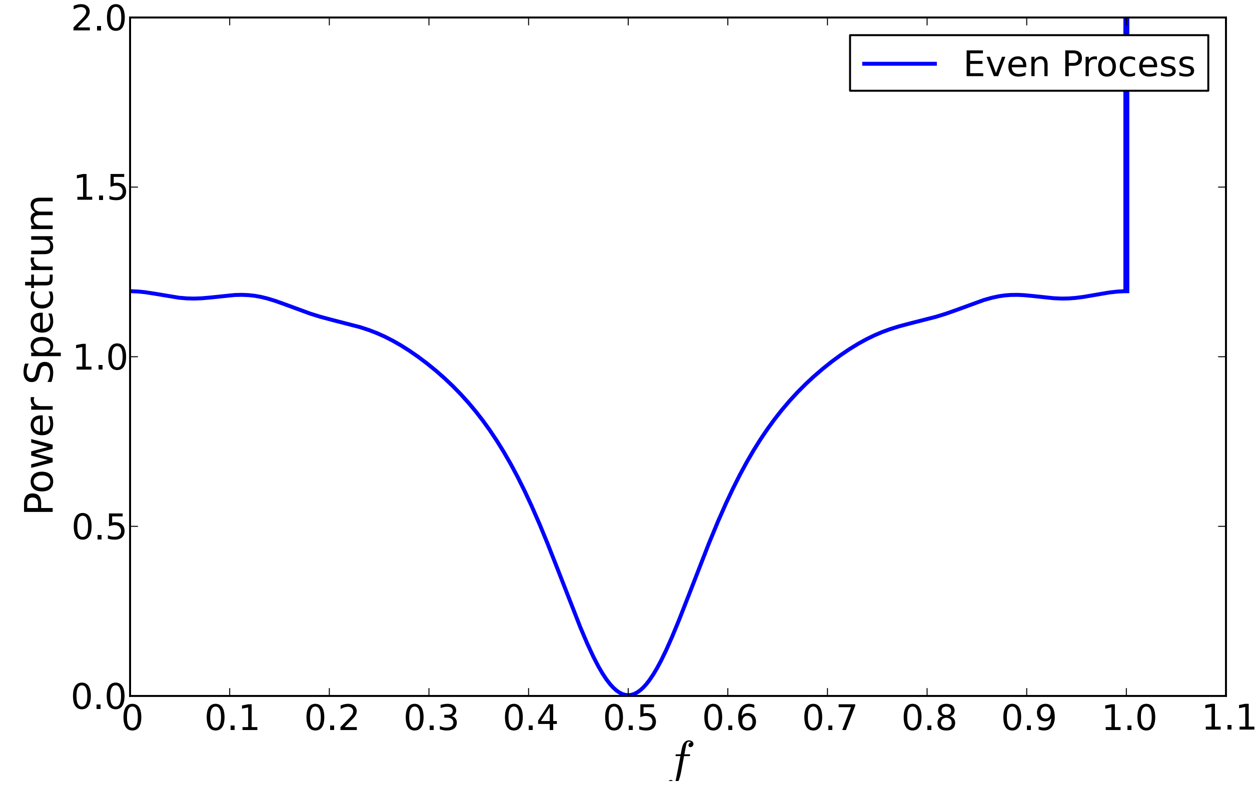


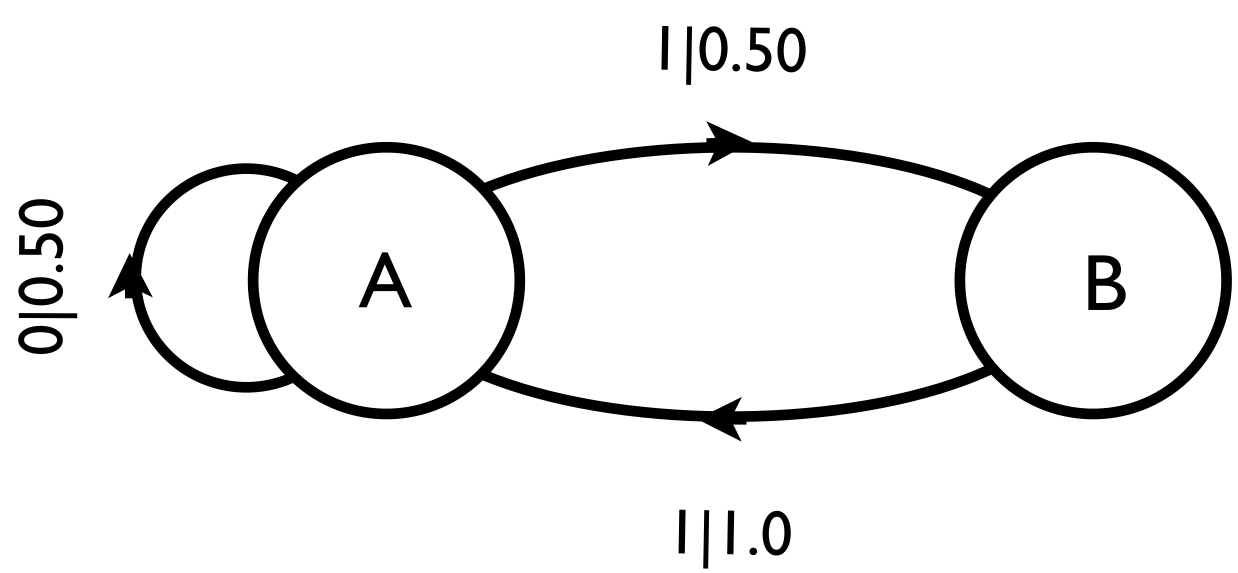




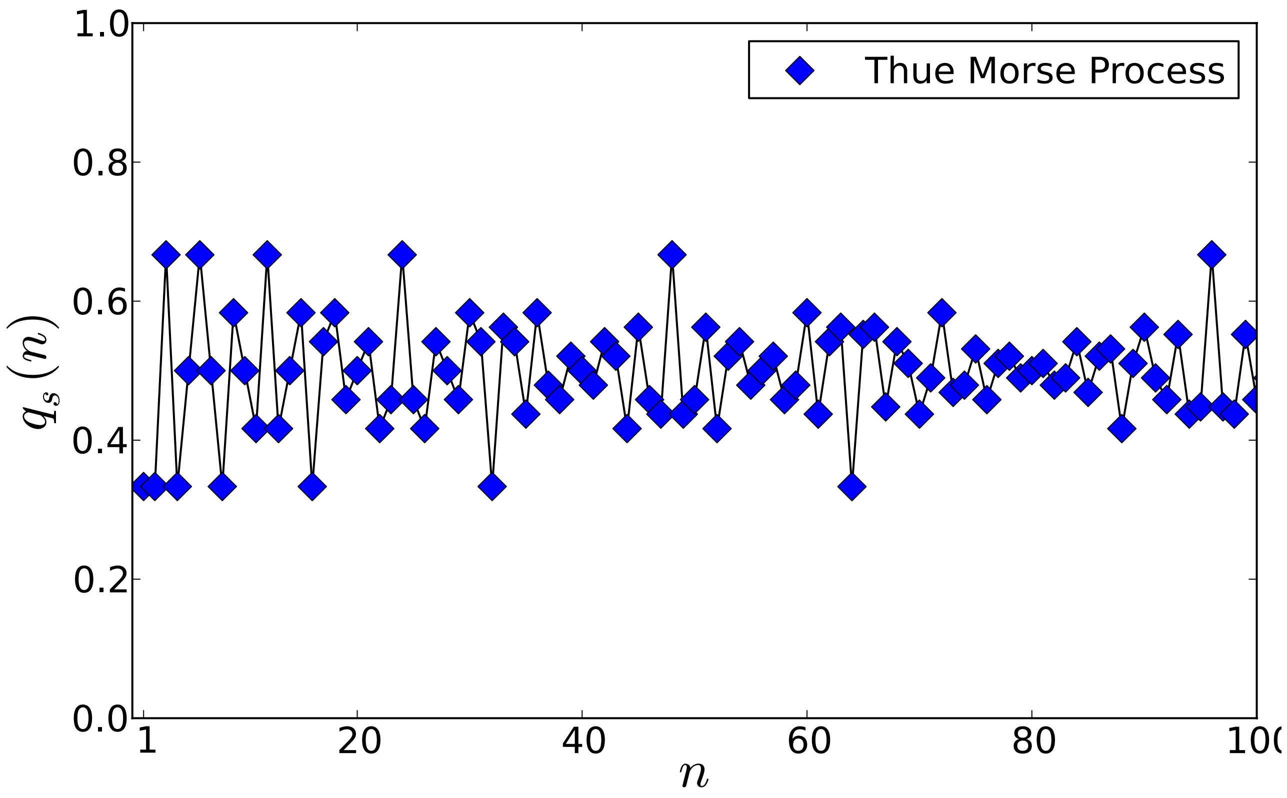
Examples of Power Spectra: Even System

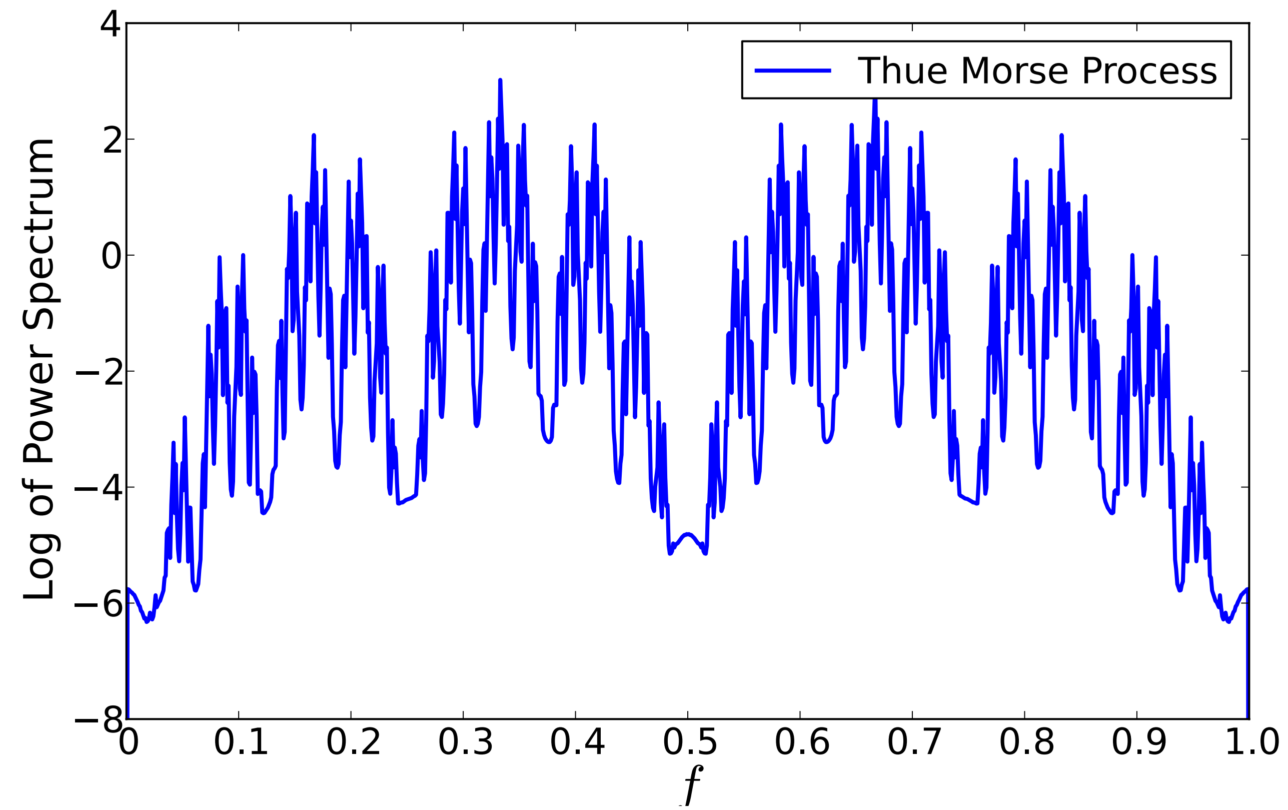






Examples of Power Spectra: Morse-Thue





We have the mapping: {0, 1} ➔ {01, 10}

Review

• Power spectra are naturally related to a two-point correlation function of the original sequence. (*Wiener-Kninchin theorem)*
• Thus, power spectra are insensitive to higher-order correlations.
• This is because finding the magnitude of the Fourier Transform throws away phase information.
• Power Spectra can be classified by their scaling behavior with the sequence size.
• They come in three types: pure point, continuous, singular continuous. A spectrum may have combinations of these three.
• Looking at the power spectrum can tell much about the statistics of the sequence.

Question:

Is it possible to infer the statistics of the sequence from knowledge of the power spectrum alone?

ε-Machine Spectral Reconstruction (εMSR)

In the standard approach to pattern discovery of a sequence, one finds the frequency of all subsequences (words) of length *r* and builds a parse tree. Histories with equivalent futures are merged to form the causal states. Since, we can find the two point correlation function from the power spectrum, perhaps we can relate these to estimate sequence probabilities.

This is what ε-machine spectral reconstruction (εMSR) does.

We begin by noting that there are constraints among the sequence probabilities:

$$\Pr(u) = \Pr(0u) + \Pr(1u) = \Pr(u0) + \Pr(u1)$$

Additionally, we require that sum of the probability of finding sequences of a given length be unity (normalization):

$$\sum_{\omega \in \mathcal{A}^{r+1}} \Pr(\omega) = 1$$

Finally, we relate correlation functions to sequence probabilities.

$$q(n) = \sum_{s=0,1} \sum_{\omega^r} \Pr(s\omega^r s)$$

for $n > 1$, and where ω^r is the subset of all sequences of length n-1.

For many spectra, the correlation functions approach an asymptotic value, and this can related to the probability of finding a 1 or 0 in the sequence by:

$$q_\infty = (\Pr(0))^2 + (\Pr(1))^2$$

We refer to these equations as *Spectral Equations* at a given *r*.

The r=1 Equations:

$$\Pr(0) = \Pr(00) + \Pr(10) = \Pr(01) + \Pr(00)$$

$$\Pr(11) + \Pr(10) + \Pr(01) + \Pr(00) = 1$$

$$q(1) = \Pr(11) + \Pr(00)$$

$$q_\infty = (\Pr(00) + \Pr(01))^2 + (\Pr(10) + \Pr(11))^2$$

We have four equations and four unknowns. We can solve!

The r=2 Equations:

$$\Pr(001) - \Pr(100) = 0 \qquad \Pr(011) - \Pr(110) = 0$$

$$\Pr(001) + \Pr(101) - \Pr(011) - \Pr(010) = 0$$

$$\Pr(111) + \Pr(101) + \Pr(011) + \Pr(001) + \Pr(110) + \Pr(100) + \Pr(010) + \Pr(000) = 1$$

$$q(1) = \Pr(111) + \Pr(110) + \Pr(000) + \Pr(001)$$

$$q(2) = \Pr(111) + \Pr(101) + \Pr(000) + \Pr(010)$$

$$\begin{aligned} q_\infty &= (\Pr(000) + \Pr(001) + \Pr(010) + \Pr(011))^2 + \\ &\quad (\Pr(100) + \Pr(101) + \Pr(110) + \Pr(111))^2, \end{aligned}$$

Eight unknowns, only **seven** equations!

Where can we get an additional constraint?

Let's use correlation function *q(3)* and write this in terms of sequences of length three.

$$q(3) = \Pr(1111) + \Pr(1101) + \Pr(1011) + \Pr(1001) + \Pr(0110) + \Pr(0010) + \Pr(0100) + \Pr(0000)$$

$$\begin{aligned} q(3) \quad = \quad & \frac{\Pr^2(111)}{\Pr(111) + \Pr(110)} + \frac{\Pr(110)\Pr(101)}{\Pr(100) + p(101)} + \\ & \frac{\Pr(101)\Pr(011)}{\Pr(010) + \Pr(011)} + \frac{\Pr(100)\Pr(001)}{\Pr(000) + \Pr(001)} + \\ & \frac{\Pr^2(000)}{\Pr(000) + \Pr(001)} + \frac{\Pr(001)\Pr(010)}{\Pr(010) + \Pr(011)} + \\ & \frac{\Pr(010)\Pr(100)}{\Pr(100) + \Pr(101)} + \frac{\Pr(011)\Pr(110)}{\Pr(111) + \Pr(110)} \end{aligned}$$

Where we have used relations of the form

$$\Pr(s_0 s_1 s_2 s_3) = \Pr(s_0 s_1 s_2)\Pr(s_3 | s_0 s_1 s_2) \approx \Pr(s_0 s_1 s_2)\Pr(s_3 | s_1 s_2) = \frac{\Pr(s_0 s_1 s_2)\Pr(s_1 s_2 s_3)}{\Pr(s_1 s_2 0) + \Pr(s_1 s_2 1)}$$

We refer to this latter approximation as *memory length reduction*.

Since we have applied memory length reduction, this effectively limits the the kinds of processes we can find to those of block Markovian of length *r*. That is, we can label all possible states by their length *r* histories.

We can then write out the εMSR algorithm as follows:

- Find the Correlation Functions from the Power Spectrum.
- Write out and solve the Spectral Equations and for sequences of a given *r*.
- We label candidate States by their length *r* histories.
- We estimate the transition probabilities between states from the sequence probabilities.
- We merge States with equivalent futures to form Causal States. This gives us a candidate ε-machine.
- We generate correlation functions and the power spectrum from the candidate ε-machine.
- We compare this with the original correlation functions and power spectrum.
- If there is insufficient agreement, we increment *r* and repeat the last six steps.

Limitations to ε-Machine Spectral Reconstruction Theory

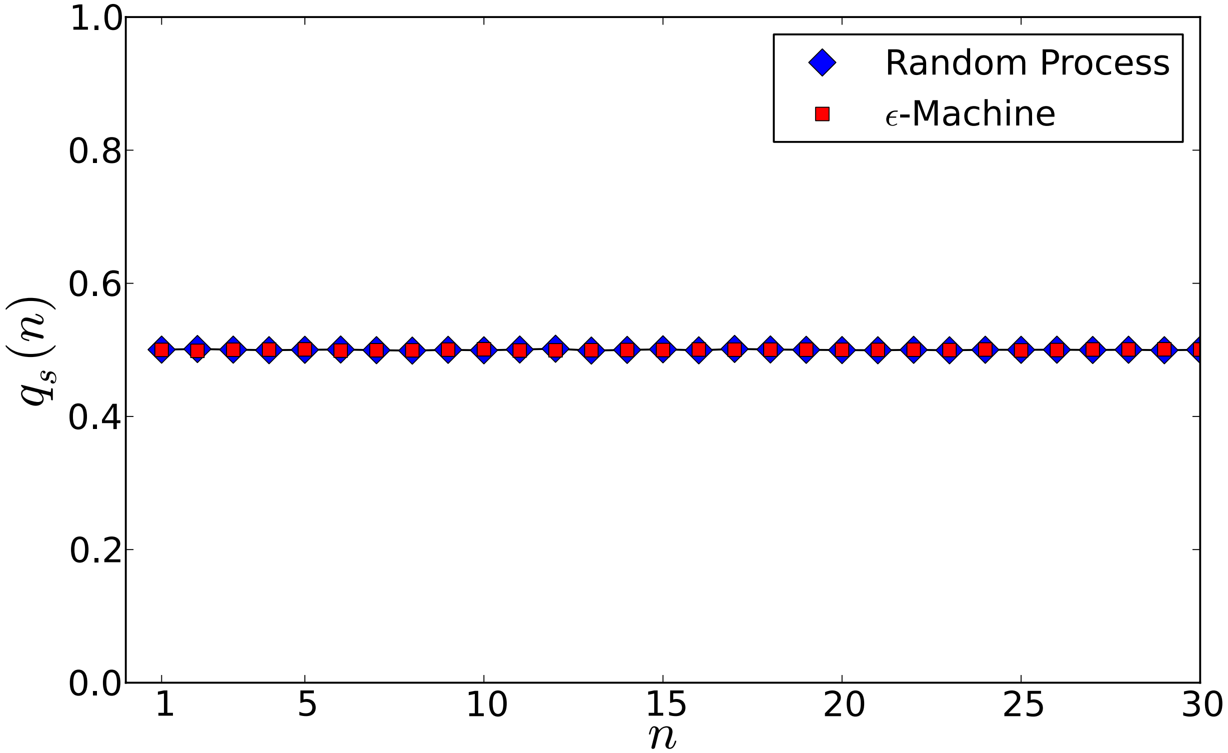
As r increases, we are forced to go to correlation functions of higher and higher n to obtain a complete set of equations. This puts a limitation on how large r can be.

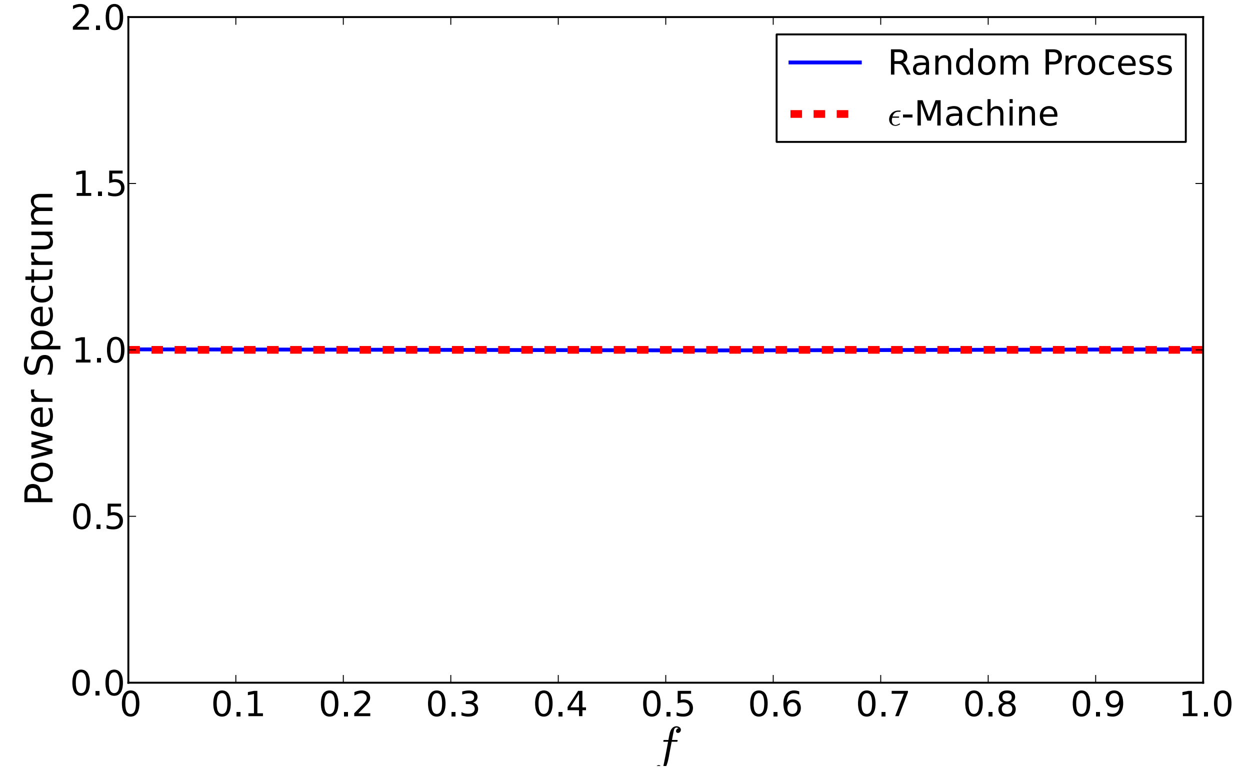
r	n_{max}	# Eqs	# Terms
2	3	8	8
3	7	16	128
4	15	32	32,768
5	31	64	$\sim 10^9$

Examples of εMSR Worked in Class

- The Random Process
- The Golden Mean Process
- The Even Process

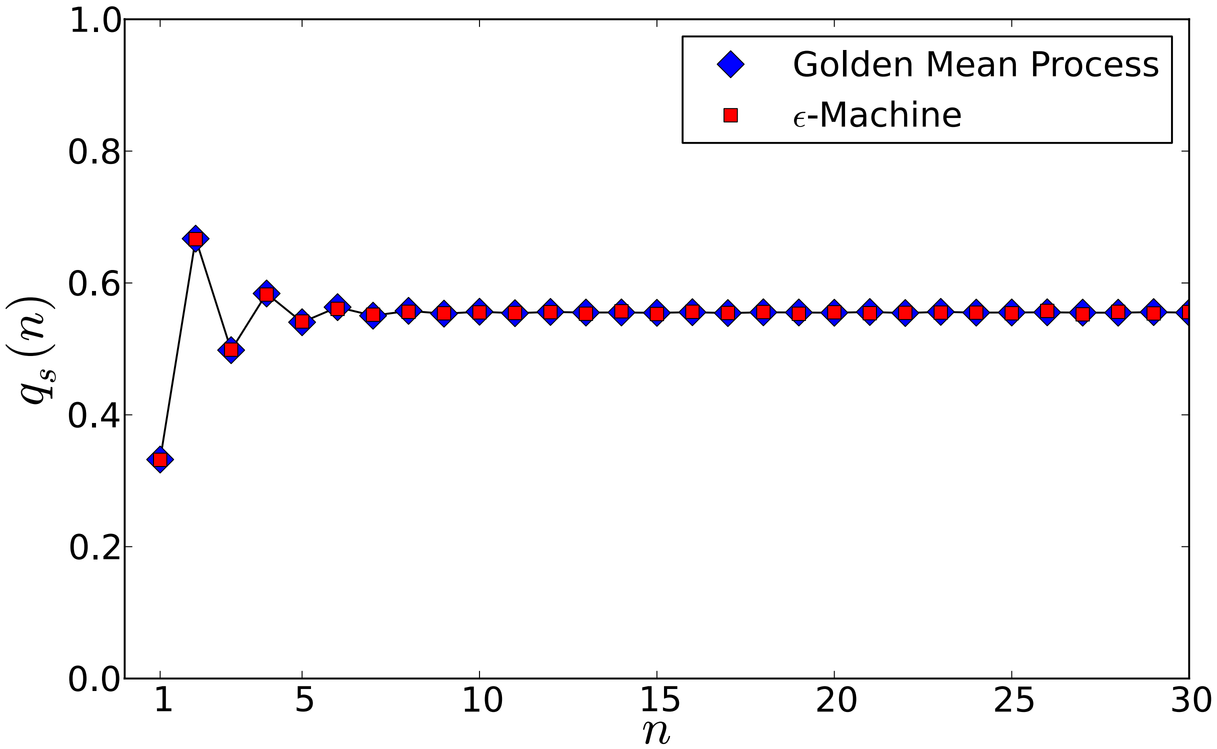
Example A: εMSR for the Random Process

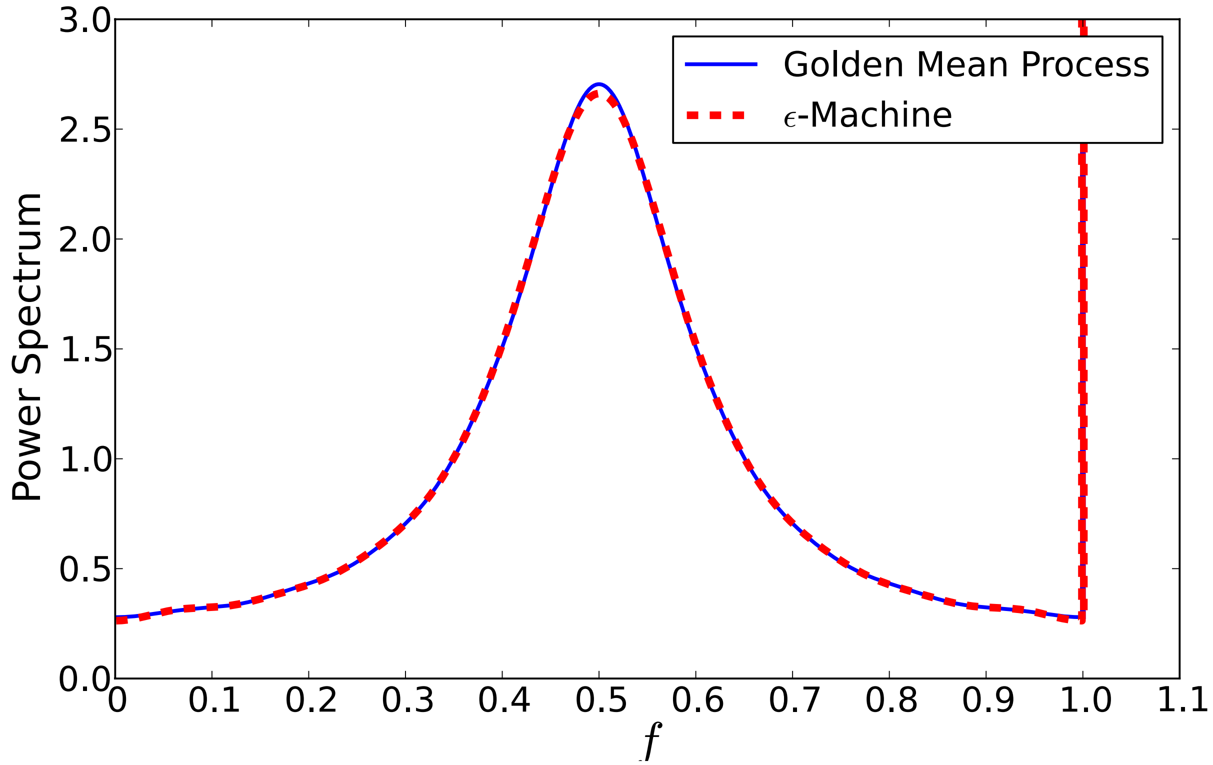




To within numerical error, εMSR reproduces the Random Process

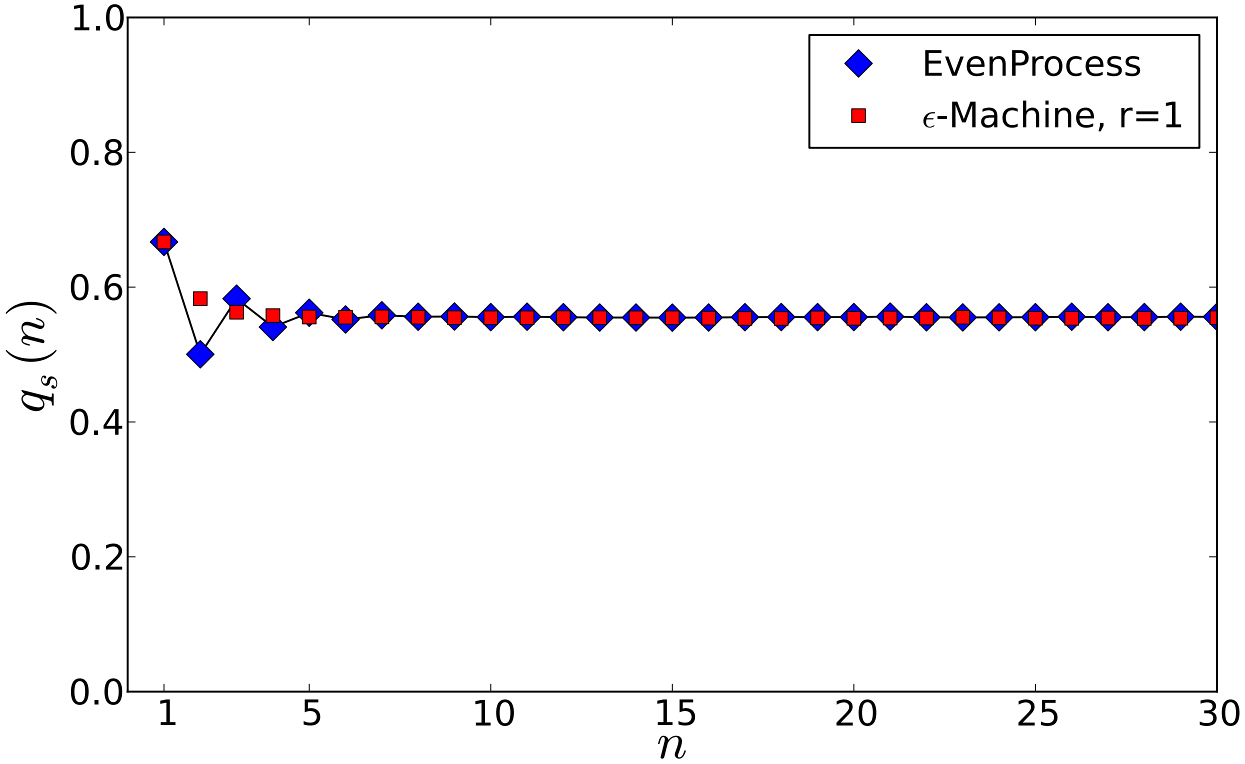
Example B: εMSR for the Golden Mean Process

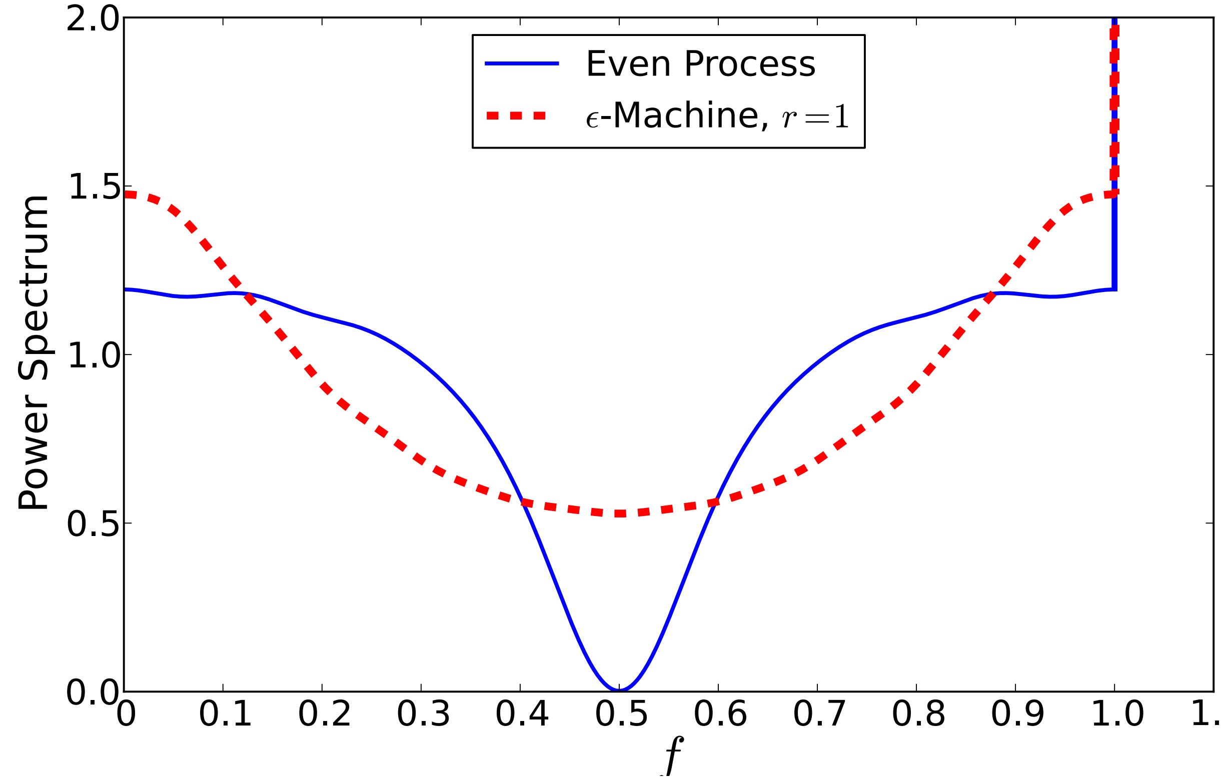


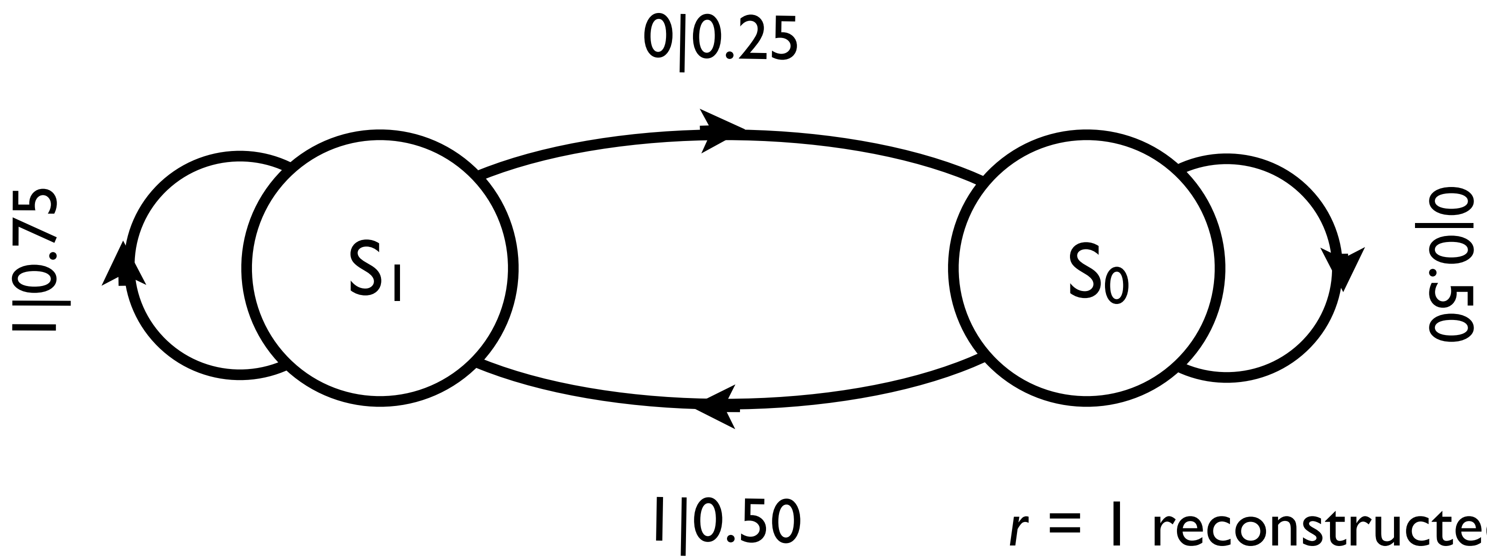


To within numerical error, εMSR reproduces the Golden Mean Process

Example C: εMSR for the Even Process, r = 1

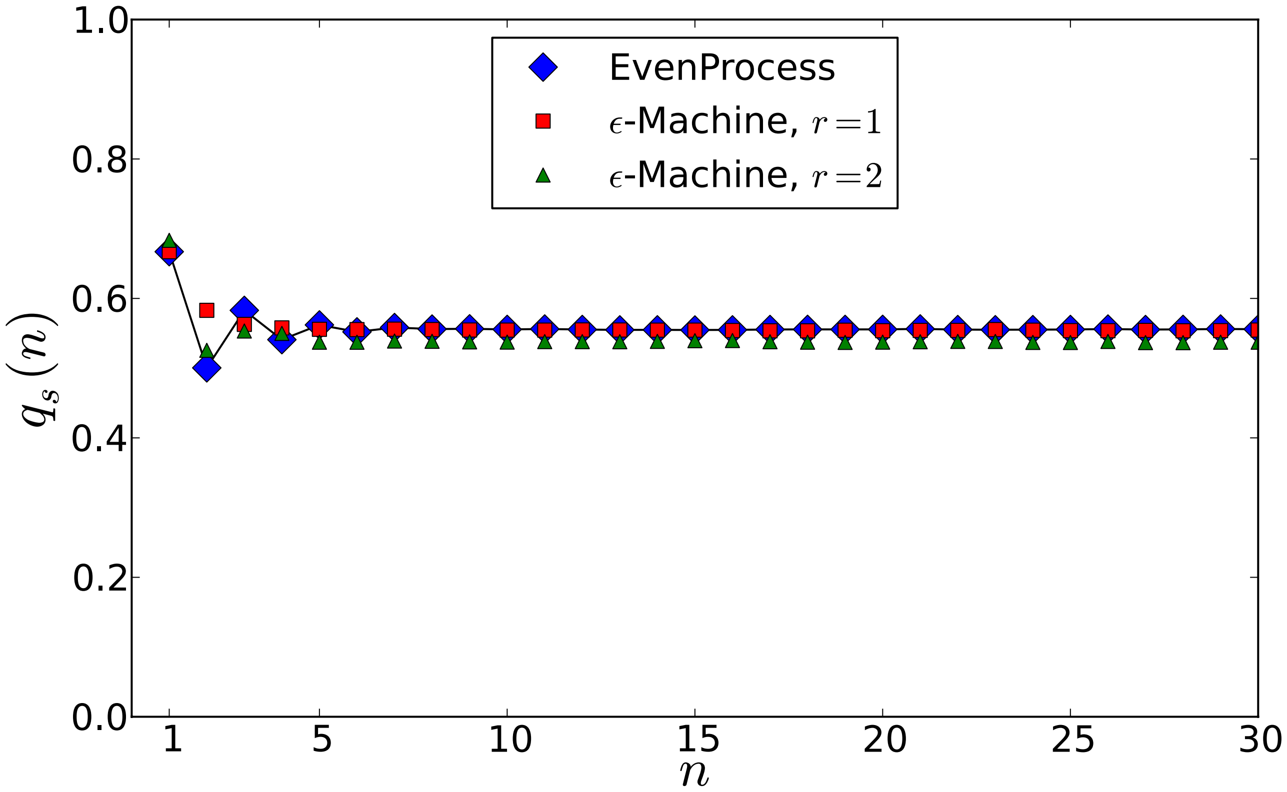


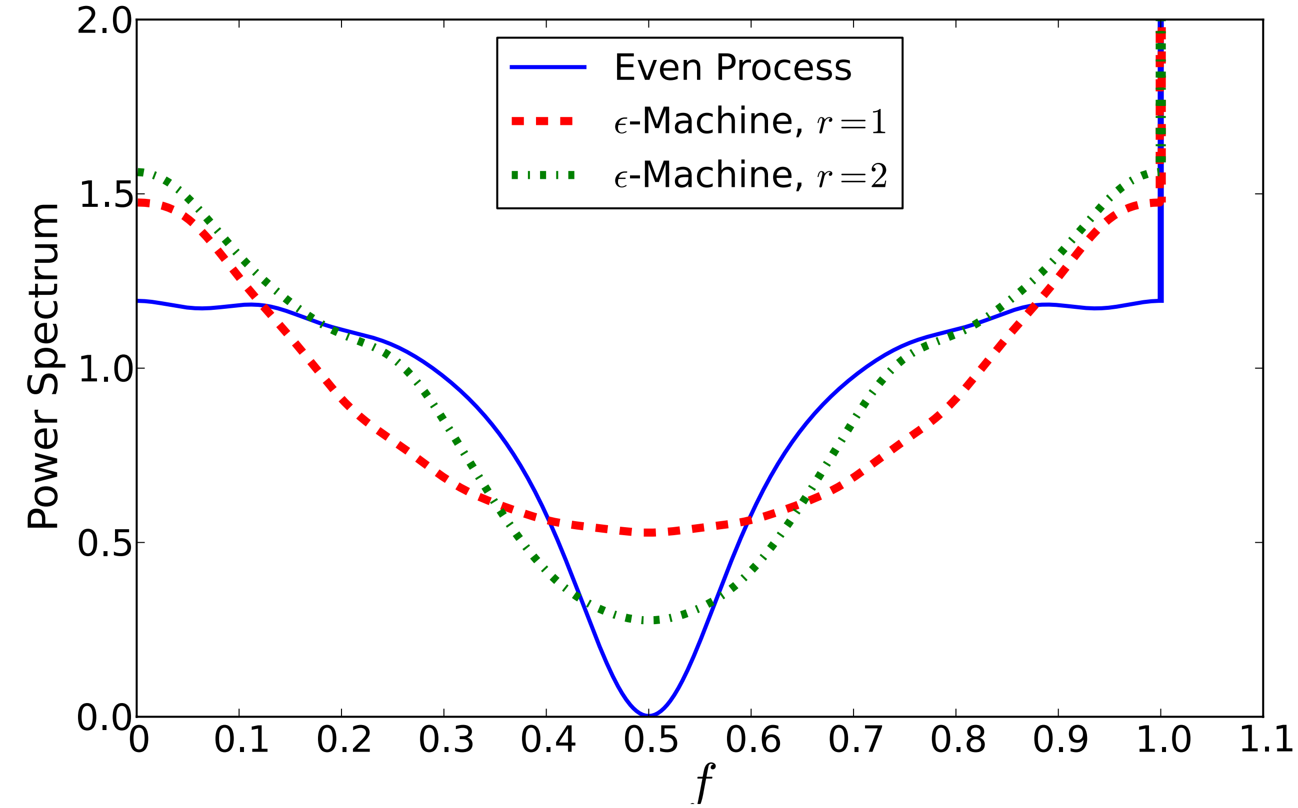


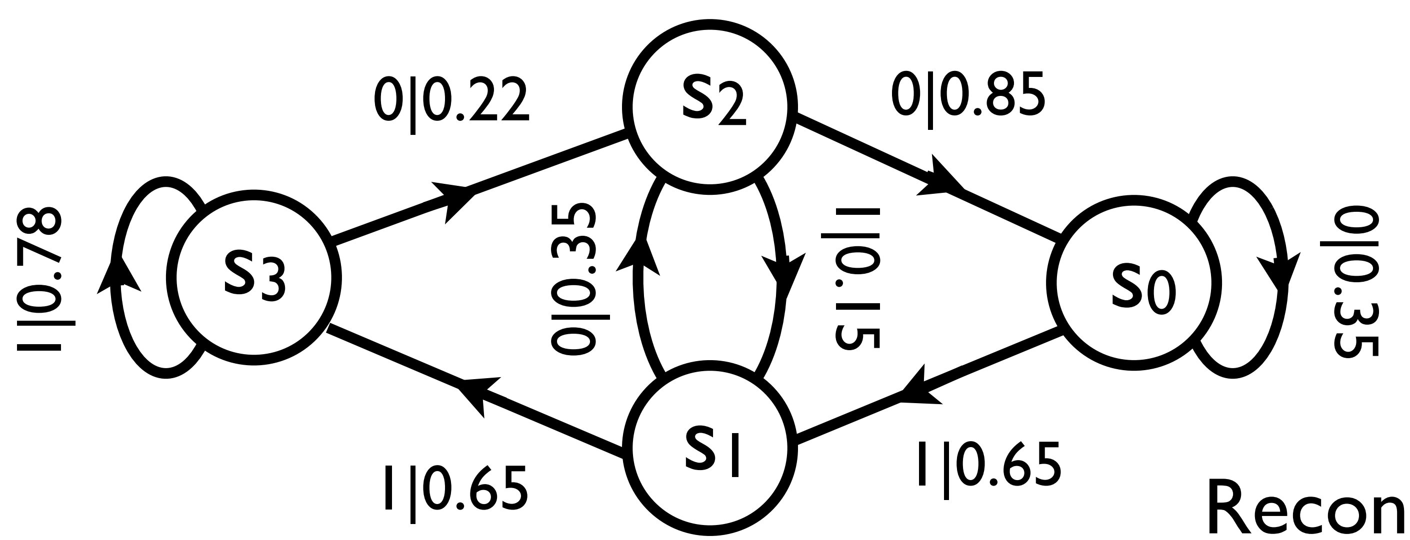


r = 1 reconstructed ε-Machine

Example C: εMSR for the Even Process, r = 2







Reconstructed r = 2
ε-machine