Lecture 27: Physics-informed deep neural networks

Professor Ilias Bilionis

Example: Sampling from the Exponential with random walk Metropolis

PREDICTIVE
SCIENCE LABORATORY

Problem definition

We want to sample from:

$$\pi(x) \propto e^{-10x}$$

using the proposal:

$$T(x_n, x) = \mathcal{N}(x|x_n, \sigma^2)$$

Let's try various sigma's.

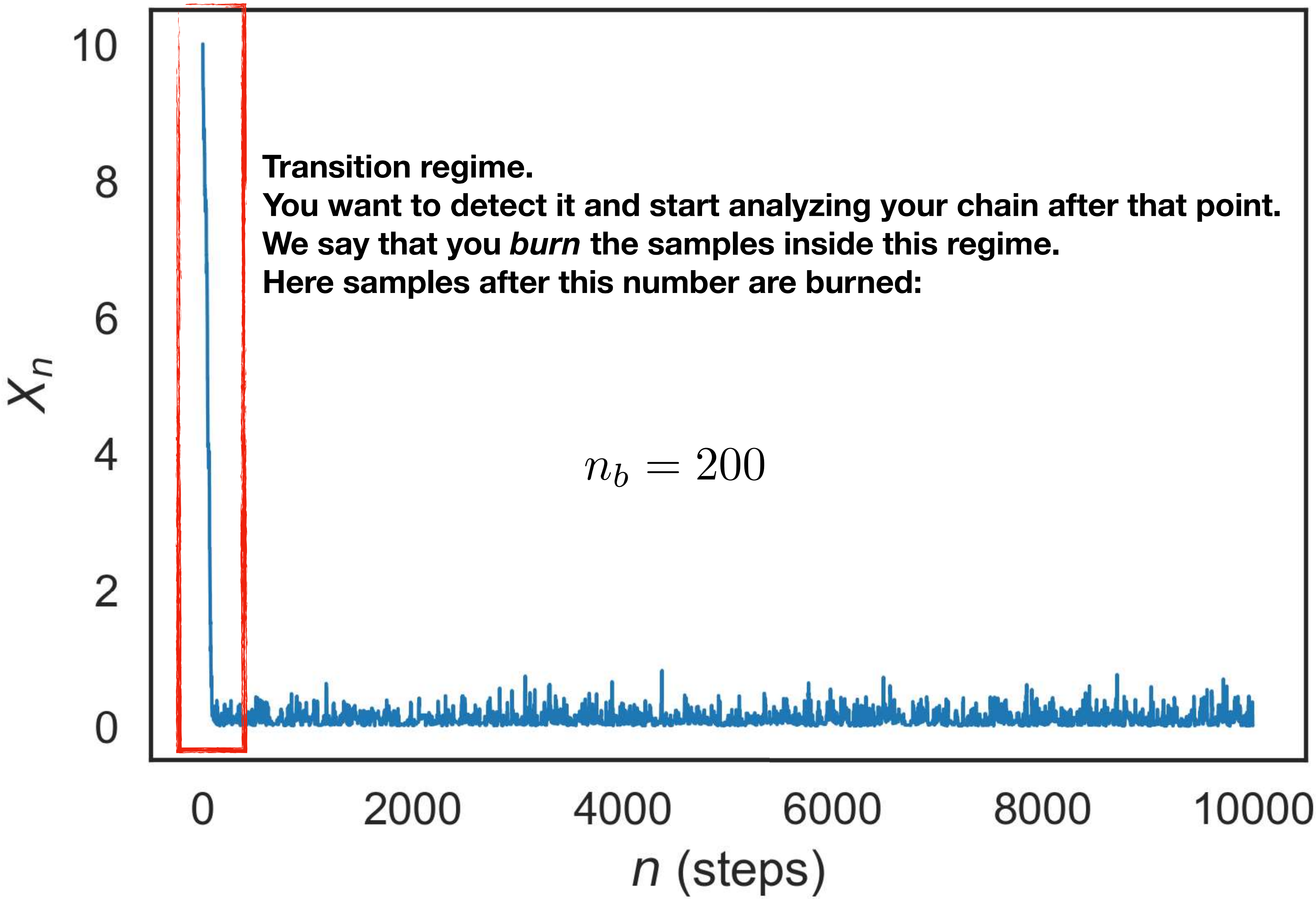
We pick the initial point to be $x_0 = 10$.

Let’s first try the “right” sigma

For sigma = 0.3, we take n=10,000 samples.

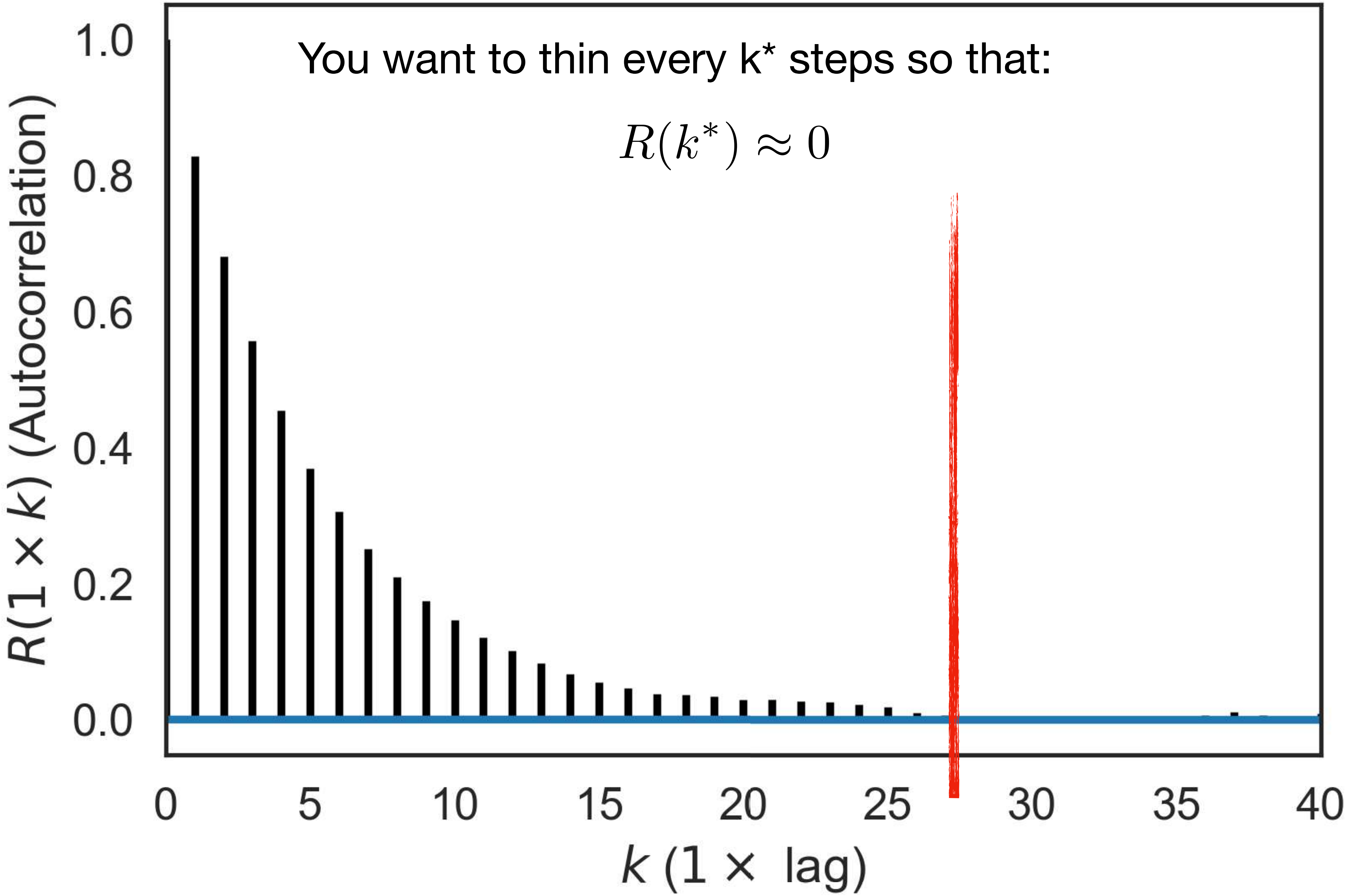
Of those, about 25% of the proposals are accepted.

Let’s visualize them.


10
8
6
4
2
0
X_n
0
2000
4000
6000
8000
10000
n (steps)
Transition regime.
You want to detect it and start analyzing your chain after that point.
We say that you *burn* the samples inside this regime.
Here samples after this number are burned:
$n_b = 200$

MCMC samples are correlated

- Ideally, you want independent samples.
- MCMC samples are correlated.
- Idea: Throw away some samples in between.
- This is called *thinning.*
- You can decide how much to thin by looking at the autocorrelation.


You want to thin every k* steps so that:
$R(k^*) \approx 0$
1.0
0.8
0.6
0.4
0.2
0.0
$R(1 \times k)$ (Autocorrelation)
0
5
10
15
20
25
30
35
40
k (1 × lag)

Autocorrelation of a Markov chain

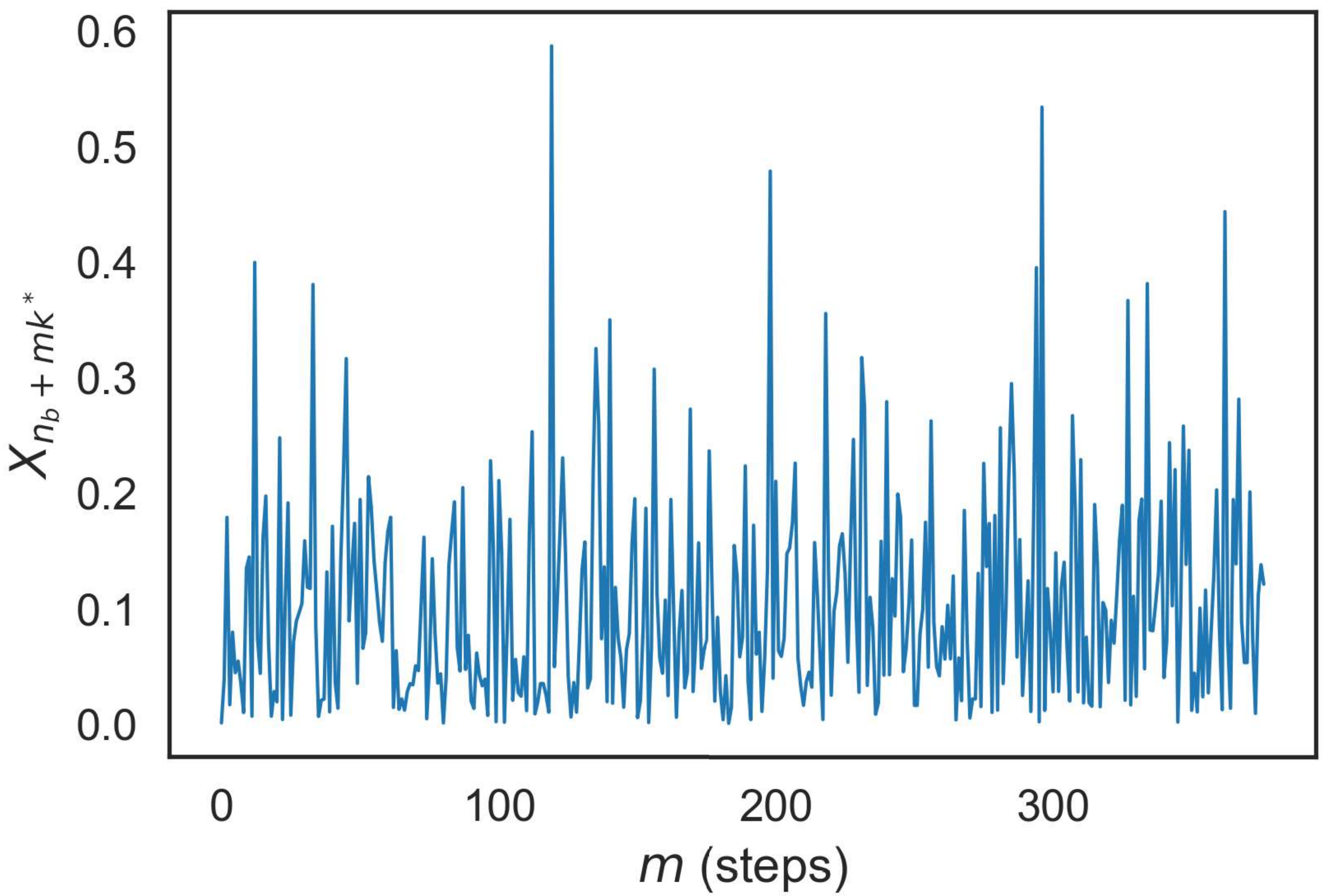
$R(k)$ = correlation between any samples of the MC that are separated by k steps

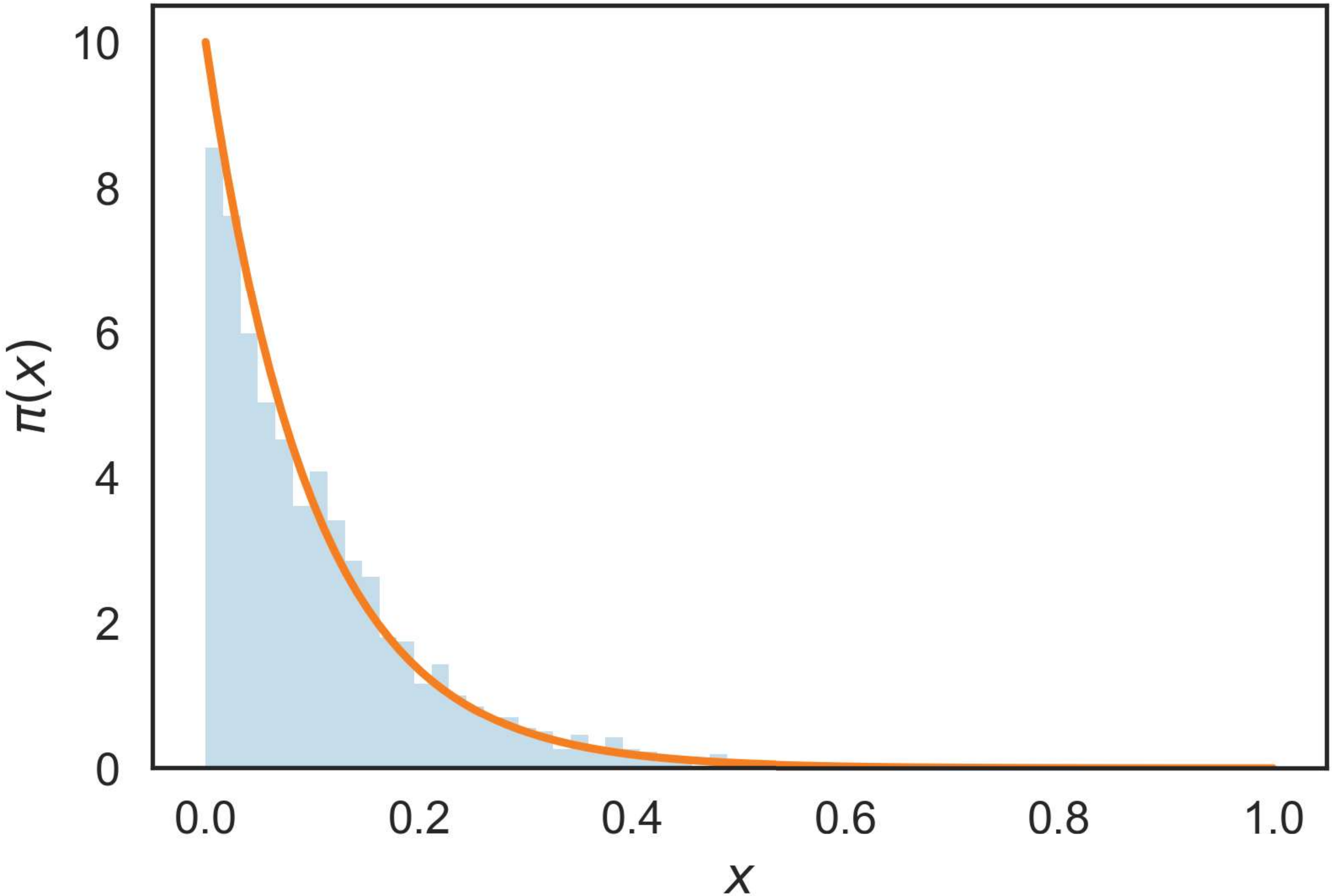
$$= \frac{\mathbb{C}[X_n, X_{n+k}]}{\mathbb{V}[X_n]}$$

PREDICTIVE
SCIENCE LABORATORY

After we burn and thin, we use in our chain:

$$X_{n_b+k^*}, X_{n_b+2k*}, \ldots$$

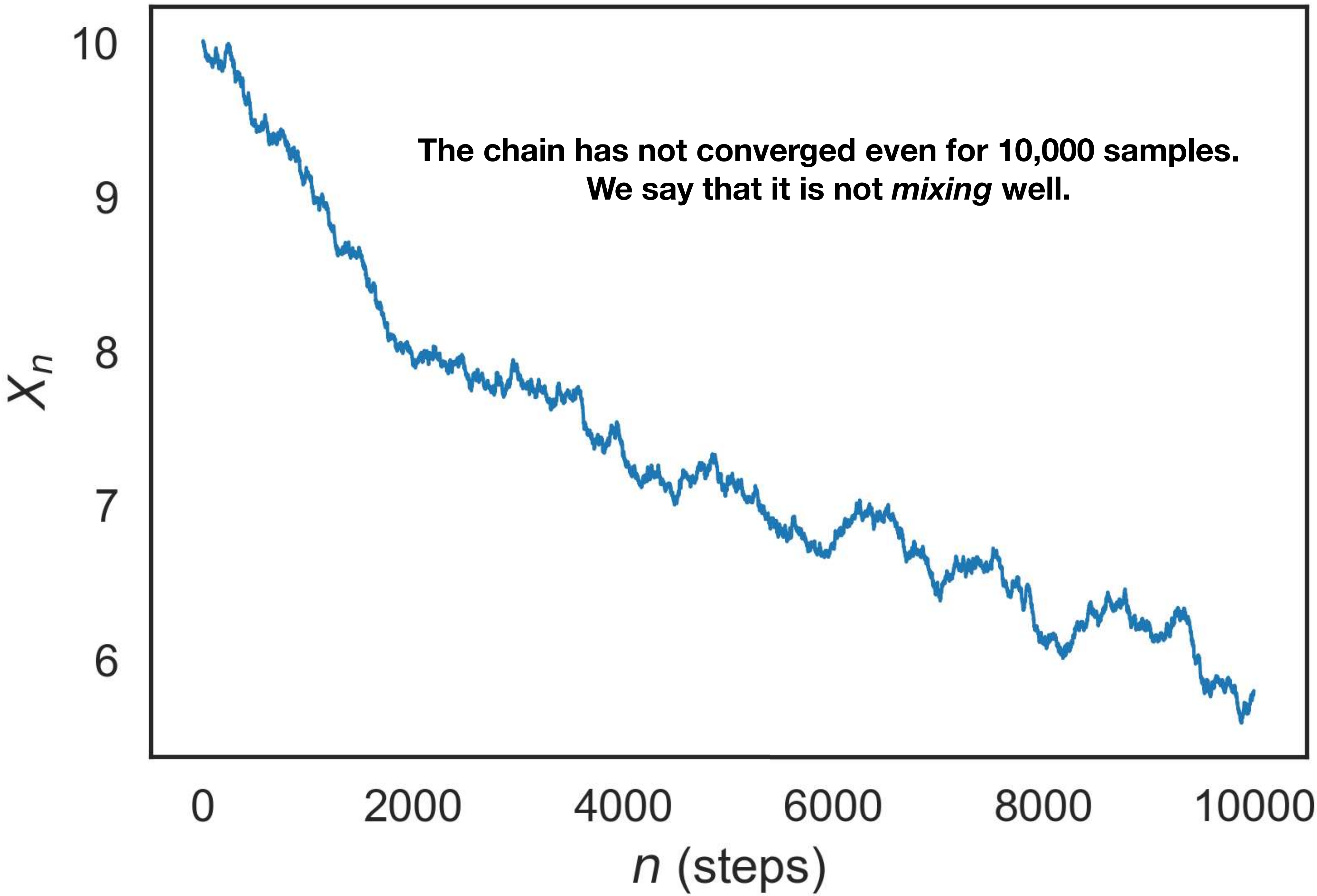




10
8
6
4
2
0
$\pi(x)$
0.0
0.2
0.4
0.6
0.8
1.0
x

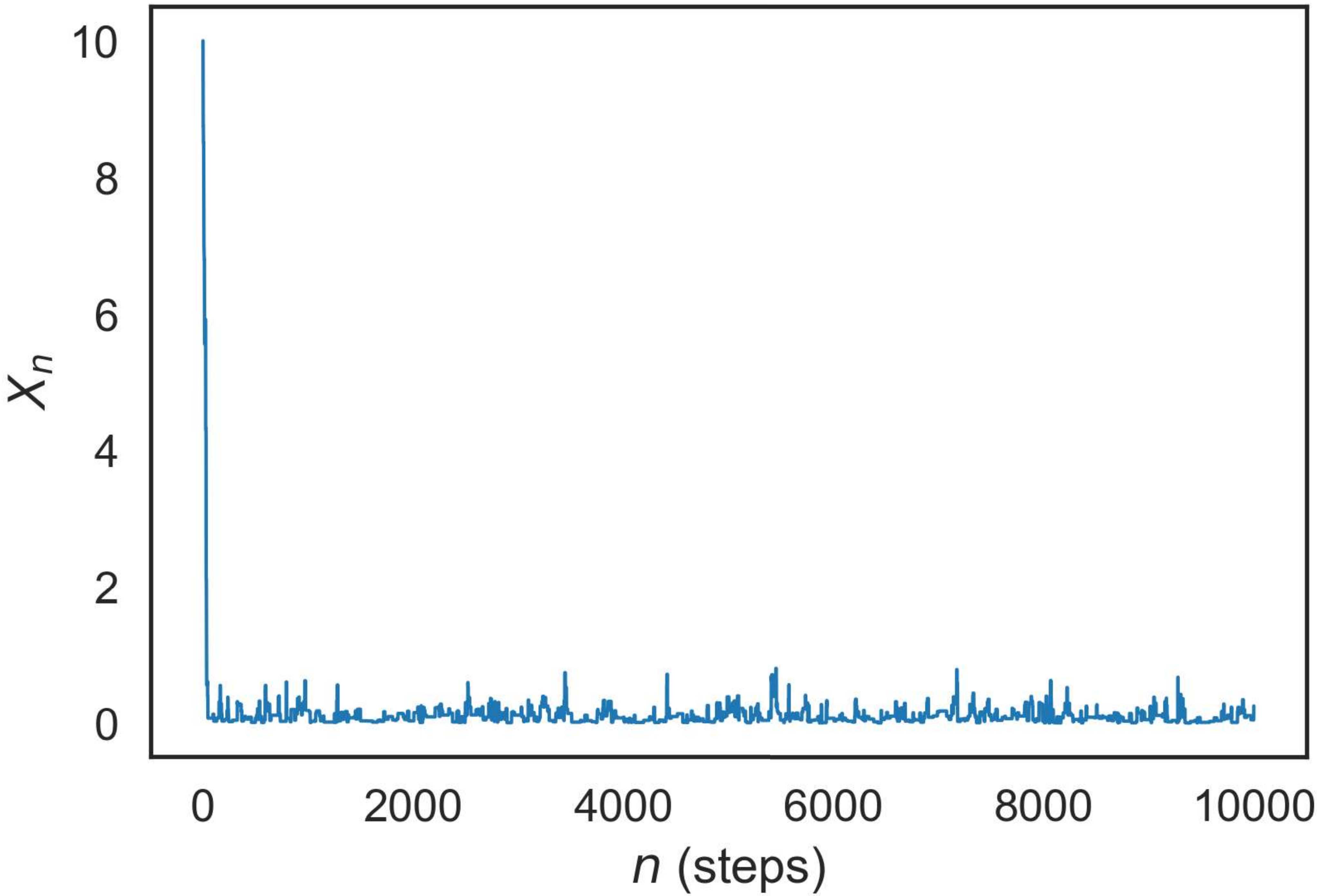
What if you pick a σ that is too small?

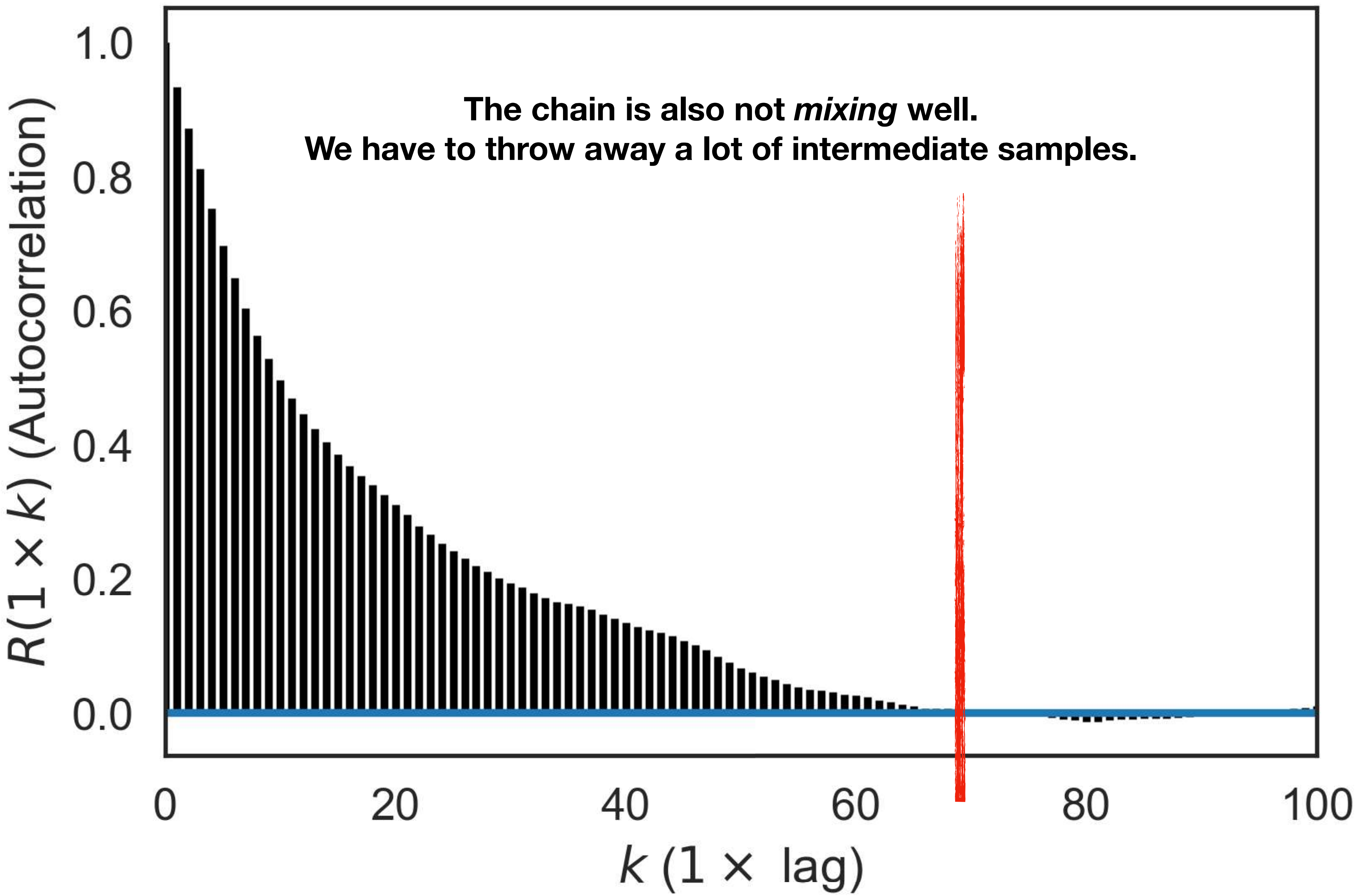
For $\sigma = 0.01$, 96% of the samples are accepted.


The chain has not converged even for 10,000 samples.
We say that it is not *mixing* well.
10
9
8
7
6
X_n
0
2000
4000
6000
8000
10000
n (steps)

What if you pick a σ that is too big?

For $\sigma = 1$, 8% of the samples are accepted.


10
8
6
4
2
0
X_n
0
2000
4000
6000
8000
10000
n (steps)


The chain is also not *mixing* well.
We have to throw away a lot of intermediate samples.
$R(1 \times k)$ (Autocorrelation)
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
k ($1 \times$ lag)

What if you pick a sigma that is too big?

For sigma = 1, 8% of the samples are accepted.

Is there a best σ?

- Yes.

- You want to keep the acceptance rate between 30% and 60% (give or take).

- You do this by *tuning* σ*.*

- $\sigma \uparrow$ *implies that acceptance rate* $\downarrow$

- $\sigma \downarrow$ *implies that acceptance rate* $\uparrow$