Tues 1-29-19
2. Probability Distribution

1. Feedback on A1
2. Descriptive statistics and probability distribution
3. Time series as realization of a process

Read notes_2.pdf

A2 due Tues, Feb 5

Feedback on A1

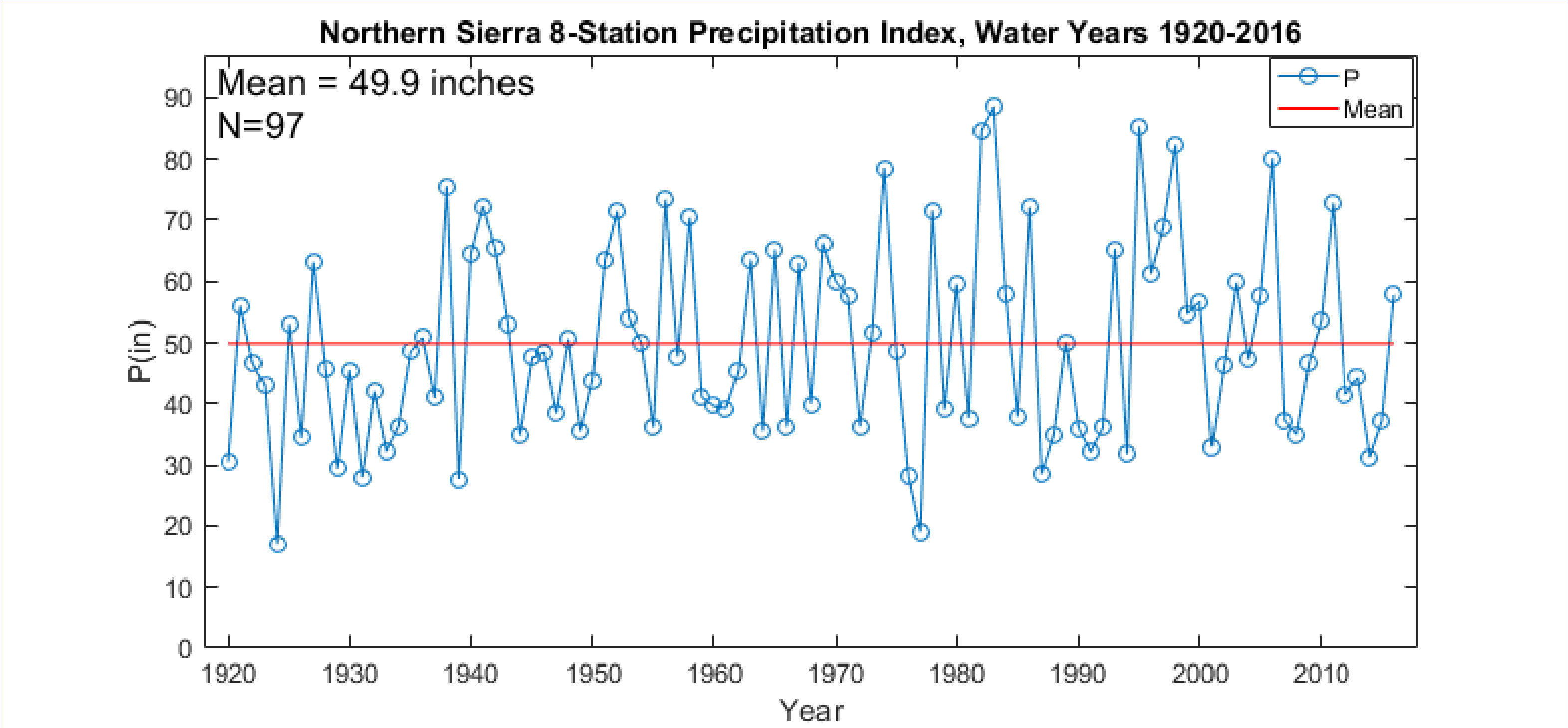
1. Download A1x.pdf from D2L
2. Automatic points, for running assignment and having uploaded by due time, is already marked in parentheses at top of first page
3. Each assignment has maximum possible 10 points; if you make no deductions, score is 10/10
4. A1x is color coded for points; purple=1; yellow=0.5; blue=0.5
5. Rubric segments are labeled in left margin by large red letters
6. Open your copy of the same assignment pdf you uploaded
7. In Acrobat Reader, using “Add text box,” mark in right margin for deductions only, with deduction and segment reference : (eg., -0.5 A)
8. At top of your pdf, mark grade like this : 9.5/10
9. If necessary, put any comments at top near the grade
10. Upload your self-graded pdf to folder A1_graded in D2L

A1 Feedback

Descriptive Statistics

- Basic statistics are mean, variance and skew
- These are measures of location, spread and symmetry of the data
- None of the 3 depend on the time sequence of observations

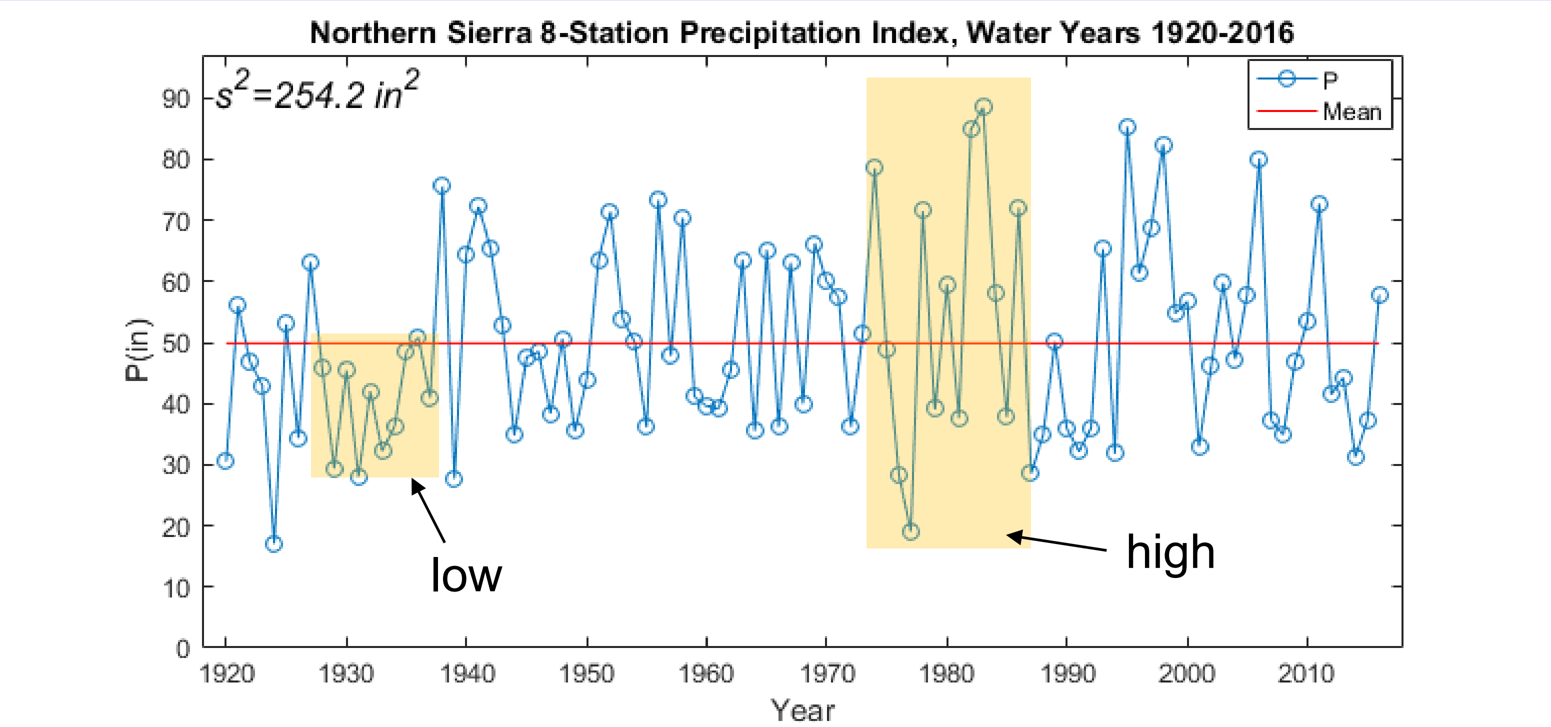
Sample Mean



- A measure of “location”
- Tends to be toward the middle, but outliers can shift location away from middle of data
- Time period of computation will affect computed statistic (sampling variability)

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$$

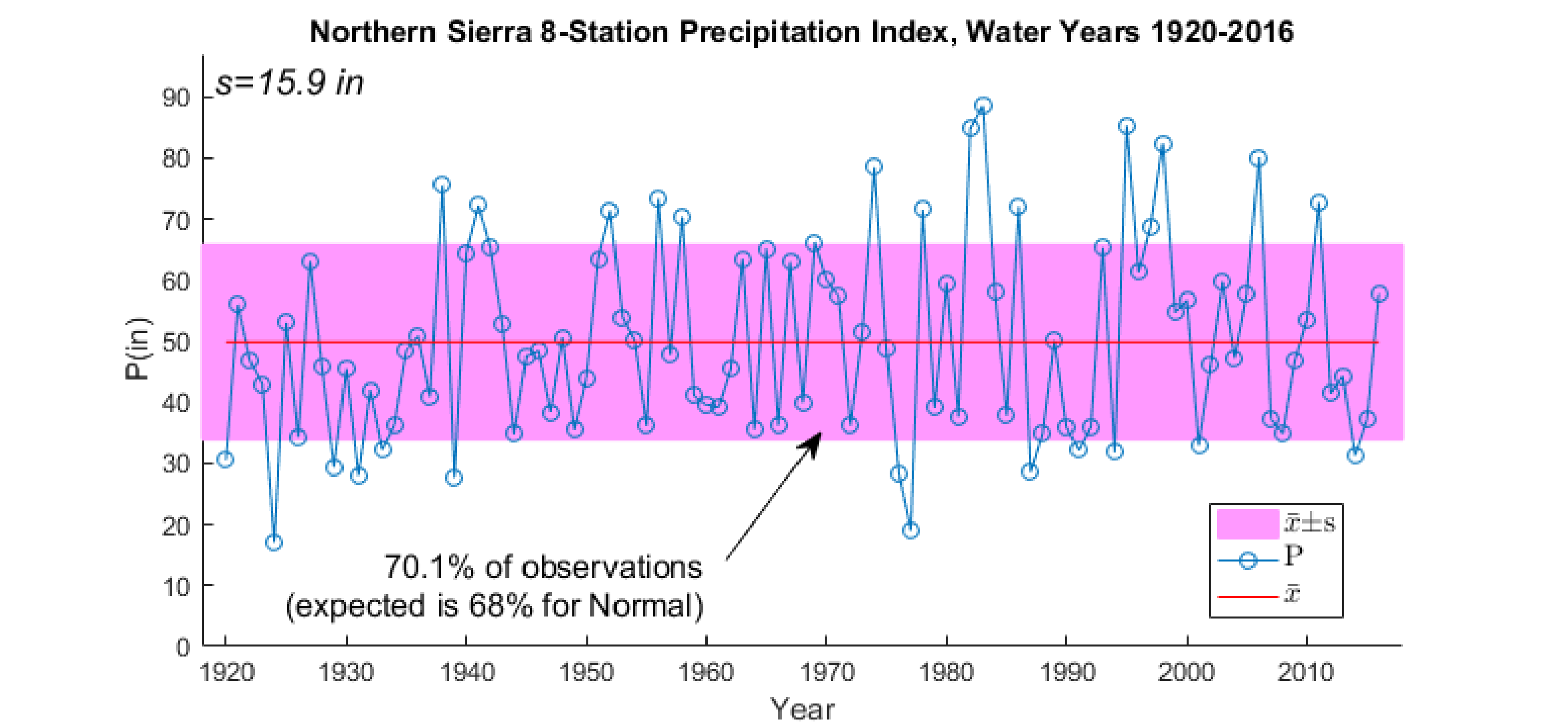
Sample Variance



- A measure of "spread"
- Average **squared** departure from **mean**
- Squaring amplifies importance of large departures, and makes sign of departure unimportant
- Units are squared data units

$$s^2 = \frac{1}{N}\sum_{t=1}^{N}(x_i - \overline{x})^2$$

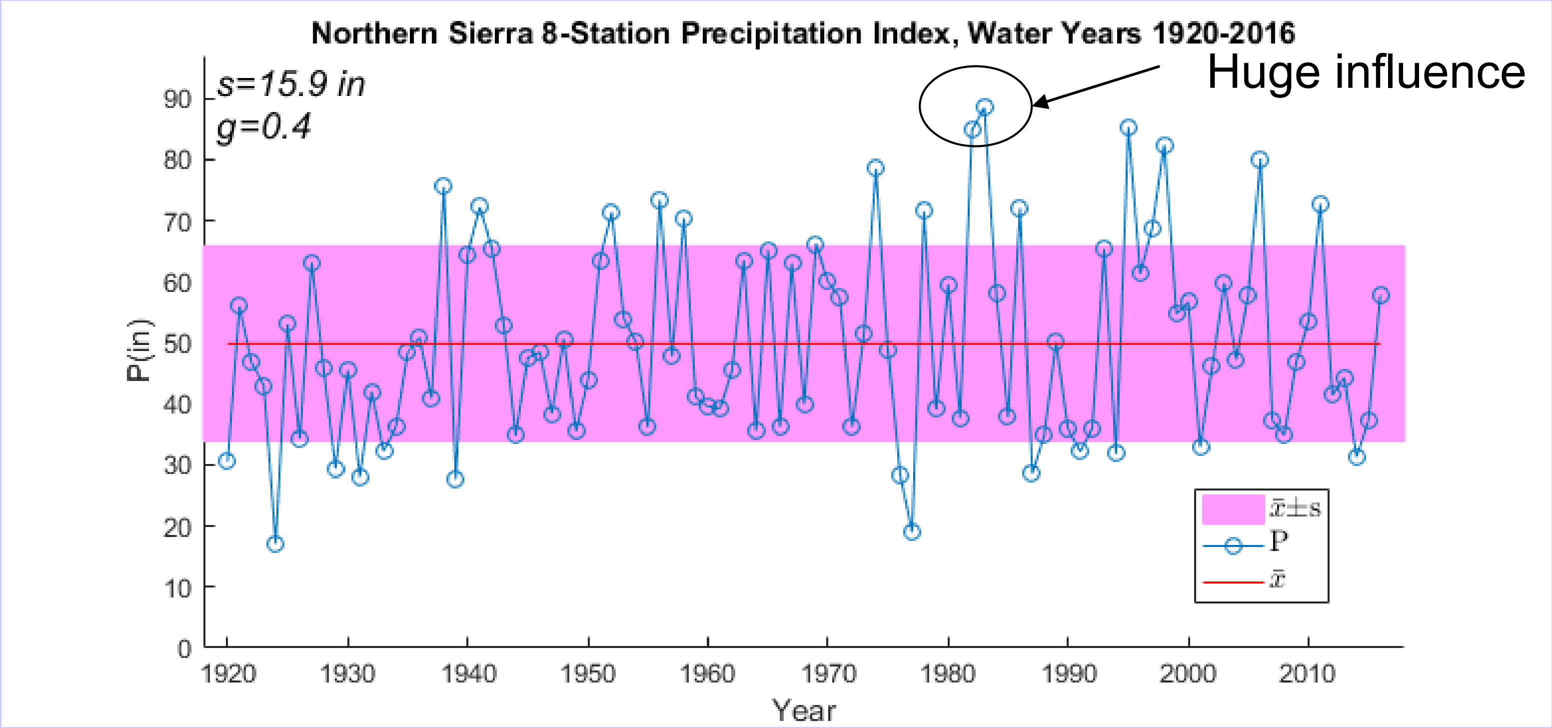
Sample Standard Deviation



- A measure of “spread” in original data units
- Square root of variance
- If data were normally distributed with same variance as *x*, would expect 68% of observations within ±*s* of mean
- Patterns of points above and below band give some idea of symmetry of distribution

$$s = \sqrt{s^2}$$

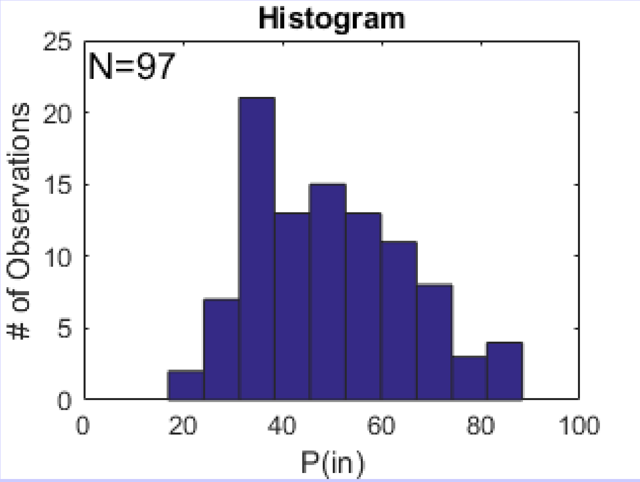
Sample Skew

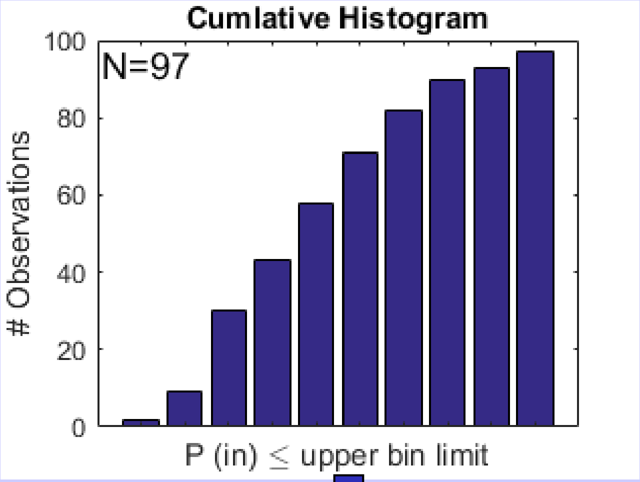


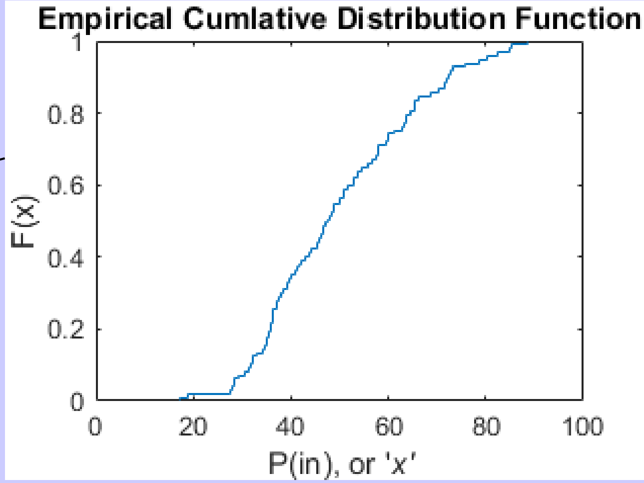
- A measure of "symmetry" of data
- Average **cubed departure** from **mean**, scaled by cube of standard deviation →skew is dimensionless
- Symmetrical data have zero skew; positive skew is skewed to right, negative to left
- Histogram is graphic

$$g = \frac{\frac{1}{N}\sum_{t=1}^{N}\left(x-\bar{x}\right)^3}{s^3}$$

Empirical Probability Distribution







"Random variable"

$$F(x) = \Pr\left[X \le x\right]$$

Random variable

Random variable: "a function that assigns real numbers to points in the sample space"

X: random variable
x: a value taken or assigned by the random variable

Usual notation

- Observed time series = realization
- Statistical generating model = process

May be a theoretical probability distribution

Normal distribution

mean

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

variance

Prob. of random variable X taking value *x*

- 2-parameter distribution
- Most commonly used distribution in time series analysis
- Plot of f(*x*) against *x* is the theoretical probability distribution (pdf)
- Plot of integral of f(*x*) against *x* is the theoretical cumulative distribution function (cdf)

disttool

Time series as realization of a process

Realization vs process

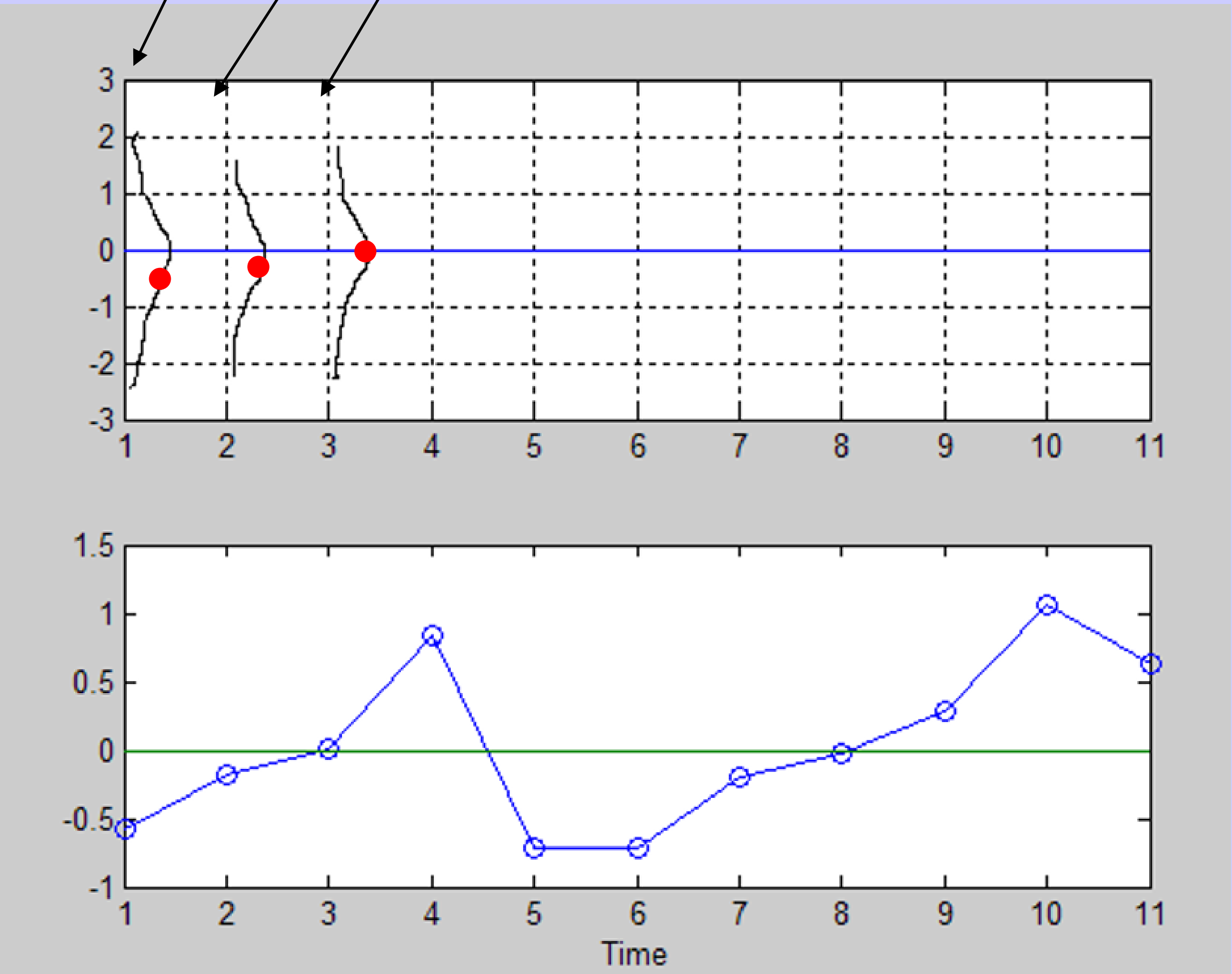
- We have a single time series --- the realization (cannot re-run time)
- Do not know the actual generation process, which can generate an infinite number of realizations
- Stuck with one realization to infer the process

Time series as realization of process

X_1 X_2 X_3

Random variables associated with times 1, 2, 3

Stochastic process



Generating Process

Observed series (realization)

Ergodic theorem

A process is called “ergodic” if the probabilistic structure of all realizations is the same. For much time series analysis, we assume ergodicity, and apply the ergodic theorem:

We can average over time and get information on the statistics of the process from a single realization

Ergodic theorem as applies to mean

Want: $\mathrm{E}\left(\mathrm{X}_t\right)$

Have $\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t$

Under ergodic assumption:

The time average $\frac{1}{N}\sum_{t=1}^{N} x_t \rightarrow \mathrm{E}\left(\mathrm{X}_t\right)$ as $N \rightarrow \infty$

(a sufficient condition for this is that the population autocorrelation dies out as lag approaches infinity --- Chatfield)