ISMT S-136 Time Series Analysis with Python

Harvard Summer School

Dmitry Kurochkin

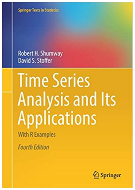
Summer 2021
Lecture 1

Contents

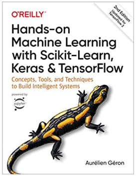
Contents

Textbooks

Required text:
Robert H. Shumway and David S. Stoffer,
Time Series Analysis and Its Applications: With R Examples,
Springer Texts in Statistics, 4th ed., 2017
ISBN: 978-3-319-52451-1
Electronic copy of the book is available at:
`https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf`



Required text:
Aurélien Géron,
Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems,
O'Reilly Media, 2nd ed., 2019
ISBN: 978-1-492-03264-9



Prerequisites

Multivariate Calculus

Multivariate calculus equivalent to MATH E-21:

- Derivatives
- Chain Rule
- Partial derivatives
- Multivariate Chain Rule
- Geometric series
- Integrals etc.

Prerequisites

Probability and Statistics

Probability and statistics equivalent to STAT E-110:

- Random variables
- Probability distributions
- Expectations
- Joint distributions
- Conditional expectations etc.

Prerequisites
Python

Python programming equivalent to CSCI E-7:

- Data types
- For loops
- If. . . elif . . . else
- Functions
- Classes/Objects etc.

Contents

Grading

Assignments, Quizzes, Exams

$$\begin{aligned}\text{Grade} = 0.25 &\cdot \text{Homework (two per week, starting June 25)}\\ + 0.20 &\cdot \text{Quizzes (two per week, starting June 23)}\\ + 0.25 &\cdot \text{Midterm (due TBA, 11:59 pm ET)}\\ + 0.30 &\cdot \text{Final (due August 6, 11:59 pm ET)}\end{aligned}$$

Contents

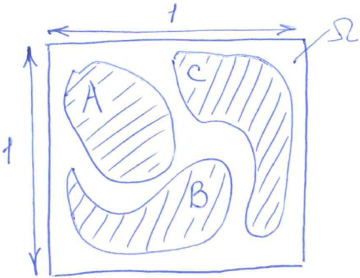
Dates of Interest

- Harvard Extension School classes begin, June 21
- Pretest is due, June 22
- Last day to change the credit status, June 23
- Course drop deadline for full-tuition refund, June 23
- Quiz 1 is due, June 23
- Assignment 1 is due, June 25
- Course drop deadline for half-tuition refund, June 30
- **Midterm Exam** is due, TBA, 11:59 pm (Eastern Time)
- Withdrawal deadline, July 23
- **Final Exam** is due, August 6, 11:59 pm (Eastern Time)

Contents

Events and Probability: Example

Ex.:



1. $P(A) = \underline{\text{area of } A} \geq 0$
2. $P(\Omega) = 1$
3. $P(A \cup B \cup C) = P(A) + P(B) + P(C)$
   if A, B, C are pairwise disjoint

Axioms of Probability

Def.

Let Ω be a set (called *sample space*) of all possible outcomes ω. Probability P is a function defined on subsets (called *events*) of Ω such that

1. For any event $A \subseteq \Omega$, $P(A) \geq 0$.
2. $P(\Omega) = 1$.
3. If events $A_1, A_2, \ldots$ are pairwise disjoint, then[1]

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots.$$

[1] Evens A and B are called *disjoint* if $A \cap B = \varnothing$.

Properties of Probability

Properties:

1. $P(\varnothing) = 0$

   Proof: Let $A_1 = \Omega, A_2 = \varnothing, A_3 = \varnothing, \ldots$, then by Axiom 3:

   $$P(\underbrace{\Omega \cup \varnothing \cup \varnothing \cup \ldots}_{=\Omega}) = P(\Omega) + P(\varnothing) + P(\varnothing) + \ldots$$

   because $\Omega, \varnothing, \varnothing, \ldots$ are pairwise disjoint.

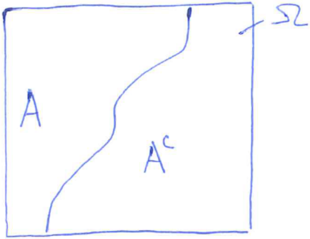
Properties of Probability

Properties (continued):

2. $P(A^c) = 1 - P(A)$.

   Proof: We notice that $\Omega = A \cup A^c$ with $A \cap A^c = \varnothing$, then

$$\underbrace{P(\underbrace{A \cup A^c}_{=\Omega})}_{=1} = P(A) + P(A^c).$$



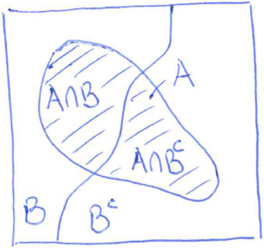
Properties of Probability

Properties (continued):

3. For any events A and B, $P(A) = P(A \cap B) + P(A \cap B^c)$.

   Proof: We notice that $A = (A \cap B) \cup (A \cap B^c)$ with $(A \cap B) \cap (A \cap B^c) = \varnothing$, then

$$P(\underbrace{(A \cap B) \cup (A \cap B^c)}_{=A}) = P(A \cap B) + P(A \cap B^c).$$



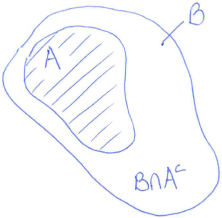
Properties of Probability

Properties (continued):

4. If $A \subseteq B$, then $P(A) \leq P(B)$.

   Proof: Notice that $B = A \cup (B \cap A^c)$ with $A \cap (B \cap A^c) = \varnothing$, then

$$P(\underbrace{A \cup (B \cap A^c)}_{B}) = P(A) + P(B \cap A^c) \geq P(A).$$



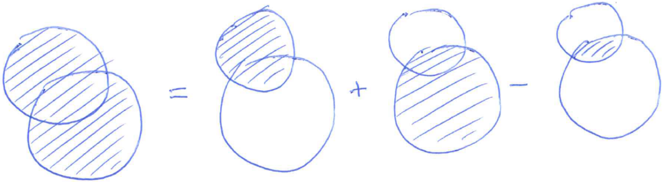
Properties of Probability

Properties (continued):

5. For any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof: We notice that $A \cup B = A \cup (B \cap A^c)$ with $A \cap (B \cap A^c) = \varnothing$, then

$$P(A \cup B) = P(A) + P(B \cap A^c) \overset{\text{by 3.}}{=} P(A) + P(B) - P(A \cap B)$$

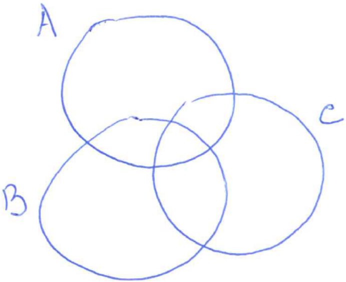


Properties of Probability

Properties (continued):

6. For any events A, B, and C,

$$
\begin{aligned}
P(A \cup B \cup C) = & P(A) + P(B) + P(C) \\
& - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
& + P(A \cap B \cap C).
\end{aligned}
$$

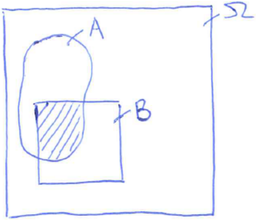


Conditional Probability: Definition

Def.

Let $B \subseteq \Omega$ be an event with $P(B) > 0$.
Conditional probability of A, given B, is defines as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$



Properties of Conditional Probability

Claim

Given $B \subseteq \Omega$ with $P(B) > 0$,
conditional probability $P(\cdot|B)$ is a probability.

Properties of Conditional Probability

Claim

Given $B \subseteq \Omega$ with $P(B) > 0$,
conditional probability $P(\cdot|B)$ is a probability.

Proof:

We just need to show that all Axioms of Probability hold:

(1) $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$ for any $A \subseteq \Omega$.

Properties of Conditional Probability

Claim

Given $B \subseteq \Omega$ with $P(B) > 0$,
conditional probability $P(\cdot|B)$ is a probability.

Proof:
We just need to show that all Axioms of Probability hold:

(1) $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$ for any $A \subseteq \Omega$.

(2) $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

Properties of Conditional Probability

Claim

Given $B \subseteq \Omega$ with $P(B) > 0$,
conditional probability $P(\cdot|B)$ is a probability.

Proof:

We just need to show that all Axioms of Probability hold:

(1) $P(A|B) = \frac{P(A\cap B)}{P(B)} \geq 0$ for any $A \subseteq \Omega$.

(2) $P(\Omega|B) = \frac{P(\Omega\cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

(3) For any pairwise disjoint events $A_1, A_2, \ldots$ we have

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \ldots |B) =& \frac{P((A_1 \cup A_2 \cup \ldots) \cap B)}{P(B)} && \text{(by def.)} \\
=& \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \ldots)}{P(B)} \\
=& \frac{P(A_1 \cap B) + P(A_2 \cap B) + \ldots}{P(B)} && \text{(by Axiom 3.)} \\
=& P(A_1|B) + P(A_2|B) + \ldots .
\end{aligned}
$$

Properties of Conditional Probability

Corollary

All properties of $P(\cdot)$ are applicable to conditional probability $P(\cdot|C)$

(assuming the conditional probability is well defined, i.e. $P(C) > 0$):

1. $P(\varnothing|C) = 0$
2. $P(A^{\mathsf{c}}|C) = 1 - P(A|C)$.
3. $P(A|C) = P(A \cap B|C) + P(A \cap B^{\mathsf{c}}|C)$.
4. If $A \subseteq B$, then $P(A|C) \leq P(B|C)$.
5. $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.

etc.

Properties of Conditional Probability

Other Properties (assuming all conditional prob. below are well defined):

1. $P(A \cap B) = P(A|B)P(B)$

Properties of Conditional Probability

Other Properties (assuming all conditional prob. below are well defined):

1. $P(A \cap B) = P(A|B)P(B)$
2. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

Properties of Conditional Probability

Other Properties (assuming all conditional prob. below are well defined):

1. $P(A \cap B) = P(A|B)P(B)$
2. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
3. $P(A) = P(A|B)P(B) + P(A|B^{\mathsf{c}})P(B^{\mathsf{c}})$

Properties of Conditional Probability

Other Properties (assuming all conditional prob. below are well defined):

1. $P(A \cap B) = P(A|B)P(B)$
2. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
3. $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$
4. $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$

Properties of Conditional Probability

Other Properties (assuming all conditional prob. below are well defined):

1. $P(A \cap B) = P(A|B)P(B)$
2. $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
3. $P(A) = P(A|B)P(B) + P(A|B^{\mathsf{c}})P(B^{\mathsf{c}})$
4. $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^{\mathsf{c}})P(A^{\mathsf{c}})}$
5. More generally:
   let $A_1, A_2, \ldots$ be partition of Ω, then

   $$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots$$

   $$P(A|B) = \frac{P(B|A)P(A)}{P(B|A_1)P(A_1)+P(B|A_2)P(A_2)+\ldots}$$

Independence: Definition

Def.

Events A and B are independent if $P(A \cap B) = P(A)P(B)$.

Independence: Definition

Def.
Events A and B are independent if $P(A \cap B) = P(A)P(B)$.

Note: If A, B are independent, then $P(A|B) = P(A)$,
given the conditional prob. is defined.

Contents

Random Variables: Definition

Def.

A real-valued function $X(\omega)$ from sample space Ω to $\mathbb{R}$ is called *random variable*.

Cumulative Distribution Function

Def.

Given a random variable X,

$$F_X(x) = P(X \leq x)$$

is called *cumultive distribution function* (cdf) of X.

Cumulative Distribution Function

Def.
Given a random variable X,

$$F_X(x) = P(X \leq x)$$

is called *cumultive distribution function* (cdf) of X.

Note: $\{X \leq x\} = \{\omega : X(\omega) \leq x\}$ is an event (i.e. a subset of Ω).

Cumulative Distribution Function

Def.

If there exists a non-negative function $f(x)$ such that for any $a \leq b$ the probability

$$P(a < X \leq b) = \int_a^b f(x)dx,$$

then $f(x)$ is called the *probability density function* (pdf).

The random variable X in that case is called *continuous random variable*.

Cumulative Distribution Function

Similarly to the cdf in the univariate case, we define the joint distribution of n random variables as follows:

Def.
Given random variable $X_1, X_2, \ldots, X_n$,

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

is called *joint cumulative distribution function* of $X_1, X_2, \ldots, X_n$.

Def.
If there exists a non-negative function $f(x_1, \ldots, x_n)$ such that for any $a_1 \le b_1, \ldots a_n \le b_n$ the probability

$$P(a_1 < X_1 \le b_1, \ldots, a_n < X_n \le b_n) = \int_{a_n}^{b_n} \ldots \int_{a_1}^{b_1} f(x_1 \ldots, x_n) dx_1 \ldots dx_n,$$

then $f(x)$ is called the *joint probability density function*.

Contents

Definition of Time Series

Def.

Time series is defined as a collection of jointly distributed random variables: $x_1, x_2, \dots, x_n$.

Def.

Time series model is a specification of the joint distribution of $x_1, x_2, \dots, x_n$.

Contents

Autocovariance Function

Def.

Given a time series x_t,

$$\gamma(s,t) = \mathrm{Cov}(x_s, x_t) = \mathrm{E}\left[(x_s - \mu_s)(x_t - \mu_t)\right]$$

is called *autocovariance function* of x_t.

Autocorrelation Function (ACF)

Def.

Given a time series x_t,

$$\begin{aligned}\rho(s,t) =& \mathrm{Corr}(x_s, x_t)\\ =& \frac{\mathrm{Cov}(x_s, x_t)}{\sqrt{\mathrm{Var}(x_s)\mathrm{Var}(x_t)}}\\ =& \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}\end{aligned}$$

is called *autocorrelation function* (ACF) of x_t.

Contents

Strict Stationarity

Def.

A time series x_t is said to be *strictly stationary* if
the following two (ordered) collections of random variables

$$x_{t_1}, x_{t_2}, \ldots, x_{t_k} \quad \text{and} \quad x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}$$

have identical joint distributions for all $k = 1, 2, \ldots$ and any choice of

$$t_1, t_2, \ldots, t_k \quad \text{and} \quad h \in \{1, 2, \ldots\}.$$

Strict Stationarity

Def.

A time series x_t is said to be *strictly stationary* if
the following two (ordered) collections of random variables

$$x_{t_1}, x_{t_2}, \ldots, x_{t_k} \quad \text{and} \quad x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}$$

have identical joint distributions for all $k = 1, 2, \ldots$ and any choice of

$$t_1, t_2, \ldots, t_k \quad \text{and} \quad h \in \{1, 2, \ldots\}.$$

Remark

If x_t is strictly stationary, then for any $h \in \{0, \pm 1, \pm 2, \ldots\}$

1. $\gamma(s+h, t+h) = \mathrm{Cov}(x_{s+h}, x_{t+h}) = \mathrm{Cov}(x_s, x_t) = \gamma(s,t) = \gamma(\underbrace{s-t}_{\text{"lag"}}, 0)$
2. $\rho(s+h, t+h) = \frac{\gamma(s+h,t+h)}{\sqrt{\gamma(s+h,s+h)\gamma(t+h,t+h)}} = \rho(s,t) = \rho(\underbrace{s-t}_{\text{"lag"}}, 0)$

(Weak) Stationarity

Def.

A time series x_t is said to be *(weakly) stationary* if

1. for all t,

$$\mu_t \doteq \mathrm{E}(x_t) = \mu,$$

i.e. does not depend on t

2. for all s, t,

$$\gamma(s,t) = \gamma(s-t,0),$$

i.e. depends on the difference $s-t$ only

Autocovariance & Autocorrelation Functions of a Stationary Time Series

If a time series x_t is stationary, we can drop the second argument, that is,

(1) $\gamma(h)$ will mean $\gamma(t+h,t) = \gamma(h,0)$
(2) $\rho(h)$ will mean $\rho(t+h,t) = \rho(h,0)$

Autocovariance & Autocorrelation Functions of a Stationary Time Series

If a time series x_t is stationary, we can drop the second argument, that is,

(1) $\gamma(h)$ will mean $\gamma(t+h,t) = \gamma(h,0)$
(2) $\rho(h)$ will mean $\rho(t+h,t) = \rho(h,0)$

We notice that

$$\begin{aligned}
\gamma(h) =& \gamma(t+h,t) = \mathrm{Cov}(x_{t+h}, x_t) = \mathrm{E}((x_{t+h}-\mu)(x_t-\mu)) \\
\rho(h) =& \rho(t+h,t) = \frac{\gamma(t+h,t)}{\sqrt{\gamma(t+h,t+h)\gamma(t,t)}} = \frac{\gamma(h)}{\gamma(0)}
\end{aligned}$$

Sample Autocovariance Function

We notice that for any choice of real numbers $a_1, a_2, \ldots, a_n \in \mathbb{R}$,

$$
\begin{aligned}
0 \leq & \operatorname{Var}\left(\sum_{j=1}^{n} a_j x_j\right) \\
= & \operatorname{Cov}\left(\sum_{j=1}^{n} a_j x_j, \sum_{k=1}^{n} a_k x_k\right) \\
= & \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \operatorname{Cov}(x_j, x_k) \\
= & \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma(j-k),
\end{aligned}
$$

i.e. $\gamma(j-k)$ is *non-negative definite*.

Sample Autocovariance Function

In order ensure non-negative results for estimated variances, we then would need to get an estimate $\hat{\gamma}(j-k)$ of the autocovariance function that is also non-negative definite:

$$0 \leq \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \hat{\gamma}(j-k).$$

Thus, we define the sample autocovariance as follows:

Def.

Given a realization $x_1, x_2, \ldots, x_n$ of a stationary time series x_t, the sample autocovariance function is defined as follows:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

for $h = 0, 1, \ldots, n-1$.

Sample Autocorrelation Function (Sample ACF)

Then the sample ACF is defined as follows:

Def.
Given $\hat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$ for a realization $x_1, x_2, \ldots, x_n$ of a stationary time series x_t, the corresponding *sample ACF* is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$