Introduction

Monday, April 3, 2023 12:57 PM

A time series is a collection of observation of random variables (x_1,x_2,....x_n) indexed by day/time in a fixed time intervals. They are not independent. Observations are just the realizations of these random variables. We have one observation for each of these variables as we take one measurement at a time. This process of collecting random variables indexed over time is called stochastic process.

E.x: observing temperature for each day.

Example of areas where time series is used:

Economics, finance: stock market, exchange rate

Social science: birth rates over time, schooling rate

Epidemiology: virus infection rate in a place during a month

Neuroscience / fMRI : check the brain cells activations

Environmental sciences: weather measurement, pollution measurements

Speech analysis: mic recording at a time.

What makes the difference between different time series?

Components:

Some time series has components and some don't. We can infer them by decomposing a timeseries. Now, let's consider

X_t = Value at a time t

We can decompose this variable into a trend (T_t) and some kind of seasonal pattern/ seasonal variation (S_t) and something that is rest (Z_t)

$$X_t = T_t+S_t+Z_t$$

Some time series has all the three. But not all .

Smoothness:

1) Some of the time series are smooth meaning, they progress steadily .

2) Some of the time series are erratic, meaning they fluctuate a lot.

Hence smoothness can quantify correlation in time. Smoothness means, the values are close to each other or pretty similar. This means the consecutive values in the series do not vary much and there is more correlation at a time.

Understanding Time series

Saturday, April 22, 2023 7:15 PM

To understand the time series

Descriptive analysis on the time series:

Using Visualizations to understand the series

Check for existence of seasonal variations, trends..

Check for smoothness to understand the correlation.

Modeling:

Fit a statistical model to have a compact description of this time series. This can enhance the our understanding of the timeseries.

We may use the model to forecast .Eg: Forecast the weather, forecast the air fare. This can help in getting prepared with the resources needed.

Time series regression:

Predicting values of one time series , by using values of the same time series or other time series in the past. That is one timeseries as a function of another timeseries

E.X: predict ozone values.

Time series of Control:

Using time signals to learn how to steer (eg: a vehicle, car...)

Stochastic dependence and Stationarity

Monday, April 3, 2023 1:00 PM

To models to time series, we need to compute statistics like mean, correlations like in regression model. What statistics do we have with time series?

Basic statistics:

We have one random variable and one observation per timestamp. Since these are actually different random variables, we cannot just average them.

Lets consider the time series example: temparature of the day

X =[x1,x2,...xn] = actual value for each day

E[X] = [ex1,ex2,ex2...exn] = Expected value for each day

Here the mean is the expected value for each day. We have one mean per day.

Marginal mean/time stamp = $\mu_x(t) = E[X_t]$

Then the actual value is the variance (how much the actual value deviates from the mean)

Marginal variance : $var(X_t)=E[(X_t - \mu_x(t))^2]$

Now we can check how do the values at different timestamps relate to each other?
Take the covariance of the random variables for each day in the standard sense of covariance and adjust the expected product of their deviations from their means.
Autocovariance is a function that gives the covariance of the process with itself at pairs of time points. Autocovariance is closely related to the autocorrelation of the process in question.

Autocovariance: $K_X(t1,t2) = cov(X_{t1}, X_{t2}) = E[(X_{t1} - \mu_x(t1))(X_{t2} - \mu_x(t2))]$

Stationarity

Saturday, April 22, 2023 7:16 PM

Stationarity:

Since the data has one mean for each timestamp, for example one data for Jan and one data for June. There is nothing we can average over. It is a poor dataset with one observation to estimate means and variance or covariance. The better way is to bring in some kind of averaging. For this, we will need to have few assumptions that actually allow us to do this averaging and they are central to essentially being able to fit the model of time series.
This concept is called Stationary.

A common assumption in many time series techniques is that the data are stationary.
A stationary process has the property that the mean, variance and autocorrelation structure do not change over time.

Weak stationarity: mean/variance is same for all X_t:

$\mu_x(t) = \mu_x,$

$\mathrm{Var}_X(t) = \sigma_x^2$

$cov(x_s, x_t) = \gamma x(s-t)$

$where\ , \gamma x = autocovariance$

Covariance is only a function of gap. i.e., cov(jan, jul)=cov(feb,aug) = ..(all months that are 6 months apart). Similarly cov(jan,feb)=cov(feb,mar)=…. (months that are 1 month apart)
Hence without having for all possible pairs, we only have one value per gap. This is the correlation and those together forms the concept of weak stationarity.

Strong stationarity: It says that distribution of a set of time variables in a time point should be same as the similar set of time variable in a different time point. i.e.;
Distribution of $X_t,\dots X_{t+n}$ = distribution of $X_{t+h}\dots . X_{t+n+h}$

Random walk method

Tuesday, April 4, 2023 7:08 PM

Consider the following random walk model. Let the starting position X_0=0, be deterministic. At each time t you flip a fair coin and add 1 to X_t-1 if you see heads H, else add -1 if you see tails T. Suppose the first 9 coin flips result in the sequence HTTTHHTTT. Compute the first 10 terms of the time series {X_t} for t=0 to 9. Find the expected position E[X_10] at time t=10

Starting from X_0 = 0, we can compute the first 10 terms of the time series {X_t} as follows:
X_0 = 0
X_1 = X_0 + 1 = 1 (since first coin flip results in H)
X_2 = X_1 - 1 = 0 (since second coin flip results in T)
X_3 = X_2 - 1 = -1 (since third coin flip results in T)
X_4 = X_3 - 1 = -2 (since fourth coin flip results in T)
X_5 = X_4 + 1 = -1 (since fifth coin flip results in H)
X_6 = X_5 + 1 = 0 (since sixth coin flip results in H)
X_7 = X_6 + 1 = 1 (since seventh coin flip results in H)
X_8 = X_7 - 1 = 0 (since eighth coin flip results in T)
X_9 = X_8 - 1 = -1 (since ninth coin flip results in T)
X_10 = X_9 - 1 = -2 (since tenth coin flip results in T)

Therefore, the first 10 terms of the time series {X_t} are:
0, 1, 0, -1, -2, -1, 0, 1, 0, -1

To find the expected position E[X_10] at time t=10, we can use the fact that the expected value of the sum of independent random variables is equal to the sum of their expected values. In this case, each coin flip is independent and has an expected value of 0, since the coin is fair and the probability of getting H or T is equal.
Therefore, we have:
E[X_10] = E[X_9 - 1] = E[X_9] - 1
Since the first 9 terms of the time series are already known, we can compute E[X_9] as the average of these values:
E[X_9] = (0 + 1 + 0 - 1 - 2 - 1 + 0 + 1 + 0) / 9 = -0.4444
Therefore, we have:
E[X_10] = E[X_9] - 1 = -0.4444 - 1 = -1.4444
So the expected position at time t=10 is approximately -1.4444.

Excercise

Saturday, April 22, 2023 7:17 PM

Find the expected position E[X_20] at time t=20

To find the expected position E[X_20] at time t=20, we can use the same approach as before. Each coin flip is independent and has an expected value of 0, so we have:
E[X_20] = E[X_19 - 1] = E[X_19] - 1
To compute E[X_19], we can use the fact that the expected value of the sum of independent random variables is equal to the sum of their expected values. Specifically, we have:
E[X_19] = E[X_18 - 1] = E[X_18] - 1
E[X_18] = E[X_17 + Y] = E[X_17] + E[Y] = E[X_17] + 0 = E[X_17]
where Y is the result of the 18th coin flip, which is independent of X_0,...,X_17 and has an expected value of 0. Therefore, we have:
E[X_19] = E[X_18] - 1 = E[X_17] - 2
Similarly, we can compute E[X_17] as:
E[X_17] = E[X_16 + Y] = E[X_16] + E[Y] = E[X_16]
where Y is the result of the 17th coin flip, which is independent of X_0,...,X_16 and has an expected value of 0. Therefore, we have:
E[X_17] = E[X_16]
Using the same reasoning, we can compute E[X_16] = E[X_15] = ... = E[X_11] = E[X_10] = -1.4444, as we found in the previous answer.
Therefore, we have:
E[X_19] = E[X_18] - 1 = E[X_17] - 2 = ... = E[X_11] - 8 = E[X_10] - 9 = -10.4444
And finally, we have:
E[X_20] = E[X_19] - 1 = -10.4444 - 1 = -11.4444
So the expected position at time t=20 is approximately -11.4444.

Find the variance of the position X_10 at time t=10
To find the variance of the position X_10 at time t=10, we can use the fact that the variance of the sum of independent random variables is equal to the sum of their variances. Specifically, we have:
Var[X_10] = Var[X_9 - 1] = Var[X_9]
To compute Var[X_9], we can use the fact that the variance of the sum of independent random variables is equal to the sum of their variances. Specifically, we have:
Var[X_9] = Var[X_8 + Y] = Var[X_8] + Var[Y]
where Y is the result of the 9th coin flip, which is independent of X_0,...,X_8 and has a variance of 1 (since it can take on the values 1 or -1 with equal probability). Therefore, we have:
Var[X_9] = Var[X_8] + 1
Similarly, we can compute Var[X_8] as:
Var[X_8] = Var[X_7 + Y] = Var[X_7] + Var[Y]
where Y is the result of the 8th coin flip, which is independent of X_0,...,X_7 and has a variance of 1.
Therefore, we have:
Var[X_8] = Var[X_7] + 1
Using the same reasoning, we can compute Var[X_7] = ... = Var[X_0] = 5.
Therefore, we have:
Var[X_9] = Var[X_8] + 1 = Var[X_7] + 2 = ... = Var[X_0] + 9 = 5 + 9 = 14
And finally, we have:
Var[X_10] = Var[X_9] = 14
So the variance of the position at time t=10 is 14.

Excersice

Saturday, April 22, 2023 7:18 PM

Find the variance of the position X_{20} at time $t = 20$
To find the variance of the position X_{20} at time $t=20$, we can use the same approach as before. Each coin flip is independent and has a variance of 1, so we have:
$Var[X_{20}] = Var[X_{19} - 1] = Var[X_{19}]$
To compute $Var[X_{19}]$, we can use the fact that the variance of the sum of independent random variables is equal to the sum of their variances. Specifically, we have:
$Var[X_{19}] = Var[X_{18} - 1] = Var[X_{18}]$
$Var[X_{18}] = Var[X_{17} + Y] = Var[X_{17}] + Var[Y] = Var[X_{17}] + 1$
where Y is the result of the 18th coin flip, which is independent of $X_0,...,X_{17}$ and has a variance of 1.
Therefore, we have:
$Var[X_{18}] = Var[X_{17}] + 1$
Similarly, we can compute $Var[X_{17}]$ as:
$Var[X_{17}] = Var[X_{16} + Y] = Var[X_{16}] + Var[Y] = Var[X_{16}] + 1$
Using the same reasoning, we can compute $Var[X_{16}] = Var[X_{15}] = ... = Var[X_{11}] = Var[X_{10}] = 14$, as we found in the previous answer.
Therefore, we have:
$Var[X_{18}] = Var[X_{17}] + 1 = Var[X_{16}] + 2 = ... = Var[X_{11}] + 7 = Var[X_{10}] + 8 = 22$
And finally, we have:
$Var[X_{19}] = Var[X_{18}] = 22$
So the variance of the position at time $t=20$ is:
$Var[X_{20}] = Var[X_{19}] = 22$
Therefore, the variance of the position at time $t=20$ is 22.

Find the forecast $E[X_{10}|X_9]$
To find the forecast $E[X_{10}|X_9]$, we can use the fact that the future values of the time series $\{X_t\}$ depend only on the current value X_9 and the random coin flip at time $t=9$. Specifically, we have:
$E[X_{10}|X_9] = E[X_9 + Y|X_9]$
where Y is the result of the 9th coin flip, which is independent of $X_0,...,X_8$ and has a distribution that is equally likely to be 1 or -1.
Therefore, we can write:
$E[X_{10}|X_9] = E[X_9] + E[Y|X_9]$
where $E[X_9]$ is the expected value of X_9 and $E[Y|X_9]$ is the expected value of Y given X_9.
Since X_9 is the sum of 9 independent coin flips that are equally likely to be 1 or -1, we have:
$E[X_9] = 9 * E[\text{coin flip}] = 9 * 0 = 0$
where E[coin flip] is the expected value of a single coin flip, which is 0 since heads and tails are equally likely.
Since Y is equally likely to be 1 or -1 given X_9, we have:
$E[Y|X_9] = P(Y=1|X_9) * 1 + P(Y=-1|X_9) * (-1)$
To compute the probabilities $P(Y=1|X_9)$ and $P(Y=-1|X_9)$, we can use the fact that the value of X_9 uniquely determines the number of heads and tails that have been observed in the first 9 coin flips.
Specifically, if $X_9 = k$, then there have been $(9+k)/2$ heads and $(9-k)/2$ tails observed so far.
Therefore, we have:
$P(Y=1|X_9) = P(\text{10th flip is heads}|X_9) = P(9+k+1 \text{ heads out of } 10+k) = (10+k-|k|)/(20-|k|)$ $P(Y=-1|X_9) = P(\text{10th flip is tails}|X_9) = P(9-k+1 \text{ heads out of } 10+k) = (10+k-|k|)/(20-|k|)$
Substituting these values, we get:
$E[Y|X_9] = [(10+k-|k|)/(20-|k|)] * 1 + [(10+k-|k|)/(20-|k|)] * (-1) = (10+k-|k| - 10 - k + |k|)/(20-|k|) = 0$
Therefore, we have:
$E[X_{10}|X_9] = E[X_9] + E[Y|X_9] = 0 + 0 = 0$
So the forecasted expected value of the position X_{10} given X_9 is 0.