Cron job

The cron command-line utility, also known as cron job is a job scheduler on Unix-like operating systems. Users who set up and maintain software environments use cron to schedule jobs (commands or shell scripts) to run periodically at fixed times, dates, or intervals. It typically automates system maintenance or administration—though its general-purpose nature makes it useful for things like downloading files from the Internet and downloading email at regular intervals. Cron is most suitable for scheduling repetitive tasks. Scheduling one-time tasks can be accomplished using the associated at utility.

Solid principle & use cases

This article is about the SOLID principles of object-oriented programming. For the .fundamental state of matter, see Solid. For other uses, see Solid (disambiguation)

In software engineering, SOLID is a mnemonic acronym for five design principles intended to make software designs more understandable, flexible, and maintainable. The principles are a subset of many principles promoted by American software engineer and instructor Robert C. Martin, first introduced in his 2000 paper Design .Principles and Design Patterns

The SOLID ideas are

The single-responsibility principle: "There should never be more than one reason for a class to change." In other words, every class should have only one responsibility

The task is divided into different members doing different things as front-end designers do design, the tester does test, and backend developer takes care of backend development part then we can say that everyone has a single job or Most of the time it happens that when programmers must add .responsibility features or new behavior, they implement everything into the existing class which is completely wrong

The open–closed principle: "Software entities ... should be open for extension, but closed for modification

Suppose developer A needs to release .extend a class behavior, without modifying it an update for a library or framework and developer B wants some modification or add some feature on that then developer B is allowed to extend the existing class created by developer A but developer B is not supposed to modify the class directly. Using this principle separates the existing code from the modified code so it provides better stability, maintainability and minimizes changes as in your code

The Liskov substitution principle: "Functions that use pointers or references to base classes must be able to use objects of derived classes without knowing it." See also design by contract

the child of a parent class should be usable in place of its parent without any
 You can understand it in a way that a farmer's son should .unexpected behavior inherit farming skills from his father and should be able to replace his father if needed. If the son wants to become a farmer, then he can replace his father but if he wants to become a cricketer then definitely the son can't replace his father even .though they both belong to the same family hierarchy

The interface segregation principle: "Many client-specific interfaces are better than one general-purpose interface

The waiter in that restaurant gave you the menu card which includes vegetarian items, non-vegetarian items, drinks, and sweets. In this case, as a customer, you should have a menu card which includes only vegetarian items, not everything which you don't eat in your food. Here the menu should be different for different types of customers. The common or general menu card for everyone can be divided into multiple cards instead of just one. Using this principle helps in reducing the side .effects and frequency of required changes

The dependency inversion principle: "Depend upon abstractions, [not] concretions

The SOLID acronym was introduced later, around 2004, by Michael Feathers

.

Although the SOLID principles apply to any object-oriented design, they can also form a core philosophy for methodologies such as agile development or adaptive software development.

Types of distributions & use cases

Bernoulli Distribution

Let's start with the easiest distribution that is Bernoulli Distribution. It is actually easier to understand than it sounds! All you cricket junkies out there! At the beginning of any cricket match, how do you decide who is going to bat or ball? A toss! It all depends on whether you win or lose the toss, right? Let's say if the toss results in a head, you win. Else, you lose. There's no midway. A **Bernoulli distribution** has only two possible outcomes, namely 1 (success) and 0 (failure), and a single trial. So the random variable X which has a Bernoulli distribution can take value 1 with the probability of success, say p, and the value 0 with the probability of failure, say q or 1-p.

Here, the occurrence of a head denotes success, and the occurrence of a tail denotes failure. Probability of getting a head = 0.5 = Probability of getting a tail since there are only two possible outcomes.

Uniform Distribution

When you roll a fair die, the outcomes are 1 to 6. The probabilities of getting these outcomes are equally likely and that is the basis of a uniform distribution. Unlike Bernoulli Distribution, all the n number of possible outcomes of a uniform distribution are equally likely.

Poisson Distribution

Suppose you work at a call center, approximately how many calls do you get in a day? It can be any number. Now, the entire number of calls at a call center in a day is modeled by Poisson distribution. Some more examples are

1. The number of emergency calls recorded at a hospital in a day.
2. The number of thefts reported in an area on a day.
3. The number of customers arriving at a salon in an hour.
4. The number of suicides reported in a particular city.
5. The number of printing errors at each page of the book.

You can now think of many examples following the same course. Poisson Distribution is applicable in situations where events occur at random points of time and space wherein our interest lies only in the number of occurrences of the event.

Exponential Distribution

Let's consider the call center example one more time. What about the interval of time between the calls ? Here, exponential distribution comes to our rescue. Exponential distribution models the interval of time between the calls.

Other examples are:

1. Length of time between metro arrivals,
2. Length of time between arrivals at a gas station
3. The life of an Air Conditioner

Exponential distribution is widely used for survival analysis. From the expected life of a machine to the expected life of a human, exponential distribution successfully delivers the result.

Central limit theory :

In probability theory, the central limit theorem (CLT) establishes that, in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution (informally a bell curve) even if the original variables themselves are not normally distributed. The theorem is a key concept in probability theory because it implies that probabilistic and statistical methods that work for normal distributions can be applicable to many problems involving other types of distributions. This theorem has seen many changes during the formal development of probability theory. Previous versions of the theorem date back to 1811, but in its modern general form, this fundamental result in probability theory was precisely stated as late as 1920, thereby serving as a bridge between classical and modern probability theory.

For example, suppose that a sample is obtained containing many observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic mean of the observed values is computed. If this procedure is performed many times, the central limit theorem says that the probability distribution of the average will closely approximate a normal distribution. A simple example of this is that if one flips a coin many times, the probability of getting a given number of heads will approach a normal distribution, with the mean equal to half the total number of flips. At the limit of an infinite number of flips, it will equal a normal distribution.

Pop book

Ex of statistics

Some examples of statistics are:

"In a recent survey of Americans, **52%** of Republicans say global warming is happening."

In this case, "52%" is a statistic, namely the percentage of Republicans in the survey sample who believe in global warming. The population is the set of all Republicans in the United States, and the population parameter being estimated is the percentage of *all* Republicans in the United States, not just those surveyed, who believe in global warming.

"The manager of a large hotel located near Disney World indicated that 20 selected guests had a mean length of stay equal to **5.6** days."

In this example, "5.6 days" is a statistic, namely the mean length of stay for our sample of 20 hotel guests. The population is the set of all guests of this hotel, and the population parameter being estimated is the mean length of stay for *all* guests.[2] Note that whether the estimator is unbiased in this case depends upon the sample selection process; see the inspection paradox.

There are a variety of functions that are used to calculate statistics. Some include:

Sample mean, sample median, and sample mode

Sample variance and sample standard deviation

Sample quantiles besides the median., quartiles and percentiles

Test statistics, such as t-statistic, chi-squared statistic, f statistic

Order statistics, including sample maximum and minimum

Sample moments and functions thereof, including kurtosis and skewness

Various functionals of the empirical distribution function

Difference between hypothesis testing & statistical test:

Statistical testing

determine whether a predictor variable has a statistically significant relationship with an outcome variable.0

estimate the difference between two or more groups.

Hypothesis testing

- Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.
- The test provides evidence concerning the plausibility of the hypothesis, given the data.
- Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

Hypothesis testing use cases

Example 1: Biology

Hypothesis tests are often used in biology to determine whether some new treatment, fertilizer, pesticide, chemical, etc. causes increased growth, stamina, immunity, etc. in plants or animals.

For example, suppose a biologist believes that a certain fertilizer will cause plants to grow more during a one-month period than they normally do, which is currently 20 inches. To test this, she applies the fertilizer to each of the plants in her laboratory for one month.

She then performs a hypothesis test using the following hypotheses:

H_0: $\mu = 20$ inches (the fertilizer will have no effect on the mean plant growth)

H_A: $\mu > 20$ inches (the fertilizer will cause mean plant growth to increase)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then she can reject the null hypothesis and conclude that the fertilizer leads to increased plant growth.

Example 2: Clinical Trials

Hypothesis tests are often used in clinical trials to determine whether some new treatment, drug, procedure, etc. causes improved outcomes in patients.

For example, suppose a doctor believes that a new drug is able to reduce blood pressure in obese patients. To test this, he may measure the blood pressure of 40 patients before and after using the new drug for one month.

He then performs a hypothesis test using the following hypotheses:

H_0: $\mu_{after} = \mu_{before}$ (the mean blood pressure is the same before and after using the drug)

HA: μafter < μbefore (the mean blood pressure is less after using the drug)

If the p-value of the test is less than some significance level (e.g. α = .05), then he can reject the null hypothesis and conclude that the new drug leads to reduced blood pressure.

Example 3: Advertising Spend

Hypothesis tests are often used in business to determine whether or not some new advertising campaign, marketing technique, etc. causes increased sales.

For example, suppose a company believes that spending more money on digital advertising leads to increased sales. To test this, the company may increase money spent on digital advertising during a two-month period and collect data to see if overall sales have increased.

They may perform a hypothesis test using the following hypotheses:

H0: μafter = μbefore (the mean sales is the same before and after spending more on advertising)

HA: μafter > μbefore (the mean sales increased after spending more on advertising)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then the company can reject the null hypothesis and conclude that increased digital advertising leads to increased sales.

Example 4: Manufacturing

Hypothesis tests are also used often in manufacturing plants to determine if some new process, technique, method, etc. causes a change in the number of defective products produced.

For example, suppose a certain manufacturing plant wants to test whether or not some new method changes the number of defective widgets produced per month, which is currently 250. To test this, they may measure the mean number of defective widgets produced before and after using the new method for one month.

They can then perform a hypothesis test using the following hypotheses:

H0: μafter = μbefore (the mean number of defective widgets is the same before and after using the new method)

HA: μafter ≠ μbefore (the mean number of defective widgets produced is different before and after using the new method)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then the plant can reject the null hypothesis and conclude that the new method leads to a change in the number of defective widgets produced per month.

Research labs in computer science

Labs

Affective Computing Group
Automatic Coordination of Teams (ACT) Lab
Center for Autonomy and AI
Center for Computer Systems Security
Center on Knowledge Graphs
Cognitive Architecture
Cognitive Learning for Vision and Robotics Lab
Collaboratory for Algorithmic Techniques and Artifical Intelligence (CATAI)
Computational Linguistics
Computational Neuroscience Lab (iLab)
Computational Social Science Laboratory
IRIS Computer Vision Lab (CV-Lab)
Data Science Lab
Database Lab (Dblab)
Haptics Robotics and Virtual
ICT Natural Language Dialogue Group
IDM Artificial Intelligence Laboratory
Information Laboratory (InfoLab)
Intelligence and Knowledge Discovery (INK) Research Lab
Integrated Media Systems Center (IMSC)
Interaction Lab
Interactive and Collaborative Autonomous Robotic Systems (ICAROS) Lab
Interactive Knowledge Capture
Machine Learning and Data Mining Lab (Melody-Lab)
Polymorphic Robotics Lab
Privacy Research Lab
Robotics and Autonomous Systems Center (RASC)
Robotic Embedded Systems Lab
Robotics Research Lab
Semantic Information Research
Speech Analysis and Interpretation Lab (SAIL)
STEEL Security Research Lab
USC Brain project
USC Center for Artificial Intelligence in Society
USC Center for Autonomy and Artificial Intelligence

Faculty

Aleksandra Korolova
Andrew Gordon
Aram Galstyan
Barath Raghavan
Bill Swartout
Bistra Dilkina
Clifford Neuman

Craig Knoblock
Cyrus Shahabi
David Kempe
David Pynadath
David Traum
Fred Morstatter
Gale Lucas
Gaurav Sukhatme
Greg Ver Steeg
Haipeng Luo
Heather Culbertson
Jay Pujara
Jelena Mirkovic
Jesse Thomason
Jiapeng Zhang
John Heidemann
Jonathan Gratch
Jonathan May
Jose Luis Ambite
Joseph Lim
Jyotirmoy Deshmukh
Kallirroi Georgila
Kristina Lerman
Laurent Itti
Leana Golubchik
Maja Mataric
Michael Zyda
Mohammad Soleymani
Muhammad Naveed
Mukund Raghothaman
Ning Wang
Paul Rosenbloom
Pedro Szekely
Ram Nevatia
Robin Jia
Satish Thittamaranahalli
Saty Raghavachary
Shaddin Dughmi
Shang-Hua Teng
Srivatsan Ravi
Stefanos Nikolaidis
Sven Koenig
Tatyana Ryutov
Ulrich Neumann
Victor Adamchik
Wei-Min Shen
Wensheng Wu
Xiang Ren

Yan Liu
Yolanda Gil

When to keep outlier or not

Outliers are unusual values in your dataset, and they can distort statistical analyses and violate their assumptions. Unfortunately, all analysts will confront outliers and be forced to make decisions about what to do with them. Given the problems they can cause, you might think that it's best to remove them from your data. But, that's not always the case. Removing outliers is legitimate only for specific reasons.

A graph that displays an outlier. Outliers can be very informative about the subject-area and data collection process. It's essential to understand how outliers occur and whether they might happen again as a normal part of the process or study area. Unfortunately, resisting the temptation to remove outliers inappropriately can be difficult. Outliers increase the variability in your data, which decreases statistical power. Consequently, excluding outliers can cause your results to become statistically significant.

Data science tools

1. SAS
2. Apache Hadoop
3. Tableau
4. TensorFlow
5. BigML
6. Knime
7. RapidMiner
8. Excel
9. Apache Flink
10. PowerBI
11. DataRobot
12. Apache Spark
13. Sap Hana
14. MongoDB
15. Python
16. Trifacta
17. Minitab
18. R
19. Apache Kafka
20. QlikView
21. MicroStrategy

22. Google Analytics
23. Julia
24. SPSS
25. MATLAB

Binary search example

Dictionary

English contains thousands of words. We cannot possibly remember every single word. So, we use dictionaries to help us find the meaning of different words. A dictionary is basically a list of words of a language in alphabetical order.

Suppose we wish to search for the meaning of the word “Serenity”. One way to find the word is to read each word of the dictionary from start and compare it to the word we want to search. But this method is very time-consuming. It will take several minutes to hours to search for the keyword.

We can optimize the task using a binary search. We go to the middle page and compare the words on that page with “Serenity”. If the “Serenity” is alphabetically smaller than the word on the middle page then we can ignore all pages on the right side. If “Serenity” is alphabetically larger than the word on the middle page then we ignore all pages on the left side. We keep repeating this process until we find the word.

Height of Students

Suppose you require some students for annual function, for some drama, or sports-related activity. The task demands someone who is taller than 5 inches.

The problem is most students aren’t sure about their height. You have a measuring instrument but measuring the height of each student is time-consuming. We can optimize the process using a binary search.

First, we will ask students to make a line in ascending order of their height. After that, we can use binary search to find all students with a height just above 5 inches.

7 vs

Volume

Volume is how much data we have – what used to be measured in Gigabytes is now measured in Zettabytes (ZB) or even Yottabytes (YB). The IoT (Internet of Things) is creating exponential growth in data. The volume of data is projected to change significantly in the coming years.

Velocity

Velocity is the speed in which data is process and becomes accessible. I remember the days of nightly batches, now if it's not real-time it's usually not fast enough.

Variety

Variety describes one of the biggest challenges of big data. It can be unstructured and it can include so many different types of data from XML to video to SMS. Organizing the data in a meaningful way is no simple task, especially when the data itself changes rapidly.

Variability

Variability is different from variety. A coffee shop may offer 6 different blends of coffee, but if you get the same blend every day and it tastes different every day, that is variability. The same is true of data, if the meaning is constantly changing it can have a huge impact on your data homogenization.

Veracity

Veracity is all about making sure the data is accurate, which requires processes to keep the bad data from accumulating in your systems. The simplest example is contacts that enter your marketing automation system with false names and inaccurate contact information. How many times have you seen Mickey Mouse in your database? It's the classic "garbage in, garbage out" challenge.

Visualization

Visualization is critical in today's world. Using charts and graphs to visualize large amounts of complex data is much more effective in conveying meaning than spreadsheets and reports chock-full of numbers and formulas.

Value

Value is the end game. After addressing volume, velocity, variety, variability, veracity, and visualization – which takes a lot of time, effort and resources – you want to be sure your organization is getting value from the data.