

Performance Evaluation and Applications

 POLITECNICO DI MILANO

Modeling Workloads

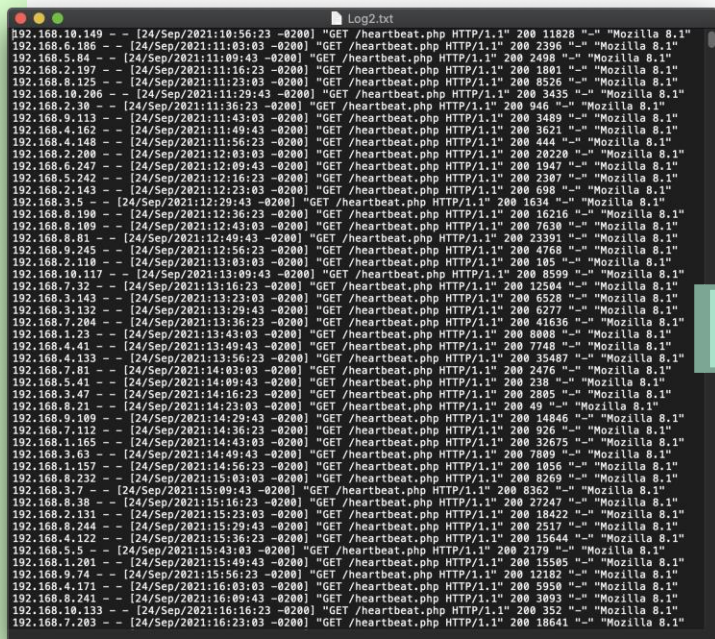
POLITECNICO DI MILANO



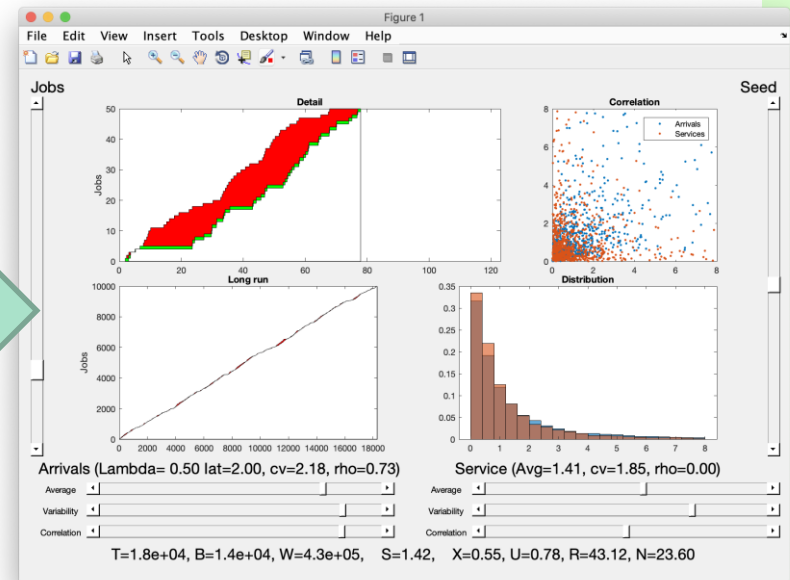
Motivating example

We have seen the importance of workload in studying the performance metrics of a system.

Starting from a set of collected measures, such as the one that can be extracted from log files, how can I quantify the workload, and properly check the amount of variability and the presence of correlation?



```
192.168.10.149 - [24/Sep/2021:10:56:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 11828 "-" "Mozilla 8.1"
192.168.8.186 - [24/Sep/2021:11:03:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 2396 "-" "Mozilla 8.1"
192.168.5.84 - [24/Sep/2021:11:09:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 2498 "-" "Mozilla 8.1"
192.168.2.197 - [24/Sep/2021:11:16:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 1801 "-" "Mozilla 8.1"
192.168.8.125 - [24/Sep/2021:11:23:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 8526 "-" "Mozilla 8.1"
192.168.10.206 - [24/Sep/2021:11:29:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 3435 "-" "Mozilla 8.1"
192.168.2.30 - [24/Sep/2021:11:36:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 946 "-" "Mozilla 8.1"
192.168.9.113 - [24/Sep/2021:11:43:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 3489 "-" "Mozilla 8.1"
192.168.4.162 - [24/Sep/2021:11:49:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 3621 "-" "Mozilla 8.1"
192.168.4.148 - [24/Sep/2021:11:56:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 444 "-" "Mozilla 8.1"
192.168.2.200 - [24/Sep/2021:12:03:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 20220 "-" "Mozilla 8.1"
192.168.6.247 - [24/Sep/2021:12:09:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 1947 "-" "Mozilla 8.1"
192.168.5.242 - [24/Sep/2021:12:16:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 2307 "-" "Mozilla 8.1"
192.168.2.143 - [24/Sep/2021:12:23:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 698 "-" "Mozilla 8.1"
192.168.3.5 - [24/Sep/2021:12:29:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 1634 "-" "Mozilla 8.1"
192.168.8.190 - [24/Sep/2021:12:36:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 16216 "-" "Mozilla 8.1"
192.168.8.109 - [24/Sep/2021:12:43:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 7630 "-" "Mozilla 8.1"
192.168.8.81 - [24/Sep/2021:12:49:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 23391 "-" "Mozilla 8.1"
192.168.9.245 - [24/Sep/2021:12:56:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 4768 "-" "Mozilla 8.1"
192.168.2.110 - [24/Sep/2021:13:03:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 105 "-" "Mozilla 8.1"
192.168.10.117 - [24/Sep/2021:13:09:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 8599 "-" "Mozilla 8.1"
192.168.7.32 - [24/Sep/2021:13:16:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 12504 "-" "Mozilla 8.1"
192.168.3.143 - [24/Sep/2021:13:23:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 6528 "-" "Mozilla 8.1"
192.168.3.132 - [24/Sep/2021:13:29:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 6277 "-" "Mozilla 8.1"
192.168.7.204 - [24/Sep/2021:13:36:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 41636 "-" "Mozilla 8.1"
192.168.1.23 - [24/Sep/2021:13:43:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 8080 "-" "Mozilla 8.1"
192.168.4.41 - [24/Sep/2021:13:49:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 7748 "-" "Mozilla 8.1"
192.168.4.133 - [24/Sep/2021:13:56:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 35487 "-" "Mozilla 8.1"
192.168.7.81 - [24/Sep/2021:14:03:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 2476 "-" "Mozilla 8.1"
192.168.5.41 - [24/Sep/2021:14:09:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 230 "-" "Mozilla 8.1"
192.168.3.47 - [24/Sep/2021:14:16:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 2805 "-" "Mozilla 8.1"
192.168.8.21 - [24/Sep/2021:14:23:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 49 "-" "Mozilla 8.1"
192.168.9.109 - [24/Sep/2021:14:29:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 14846 "-" "Mozilla 8.1"
192.168.7.112 - [24/Sep/2021:14:36:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 924 "-" "Mozilla 8.1"
192.168.1.165 - [24/Sep/2021:14:43:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 32675 "-" "Mozilla 8.1"
192.168.3.63 - [24/Sep/2021:14:49:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 7809 "-" "Mozilla 8.1"
192.168.1.157 - [24/Sep/2021:14:56:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 1856 "-" "Mozilla 8.1"
192.168.8.222 - [24/Sep/2021:15:03:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 8269 "-" "Mozilla 8.1"
192.168.3.7 - [24/Sep/2021:15:09:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 8362 "-" "Mozilla 8.1"
192.168.3.38 - [24/Sep/2021:15:16:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 27247 "-" "Mozilla 8.1"
192.168.2.131 - [24/Sep/2021:15:23:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 18422 "-" "Mozilla 8.1"
192.168.8.244 - [24/Sep/2021:15:29:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 2517 "-" "Mozilla 8.1"
192.168.4.122 - [24/Sep/2021:15:36:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 15644 "-" "Mozilla 8.1"
192.168.5.5 - [24/Sep/2021:15:43:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 2179 "-" "Mozilla 8.1"
192.168.1.201 - [24/Sep/2021:15:49:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 15505 "-" "Mozilla 8.1"
192.168.9.74 - [24/Sep/2021:15:56:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 12182 "-" "Mozilla 8.1"
192.168.4.171 - [24/Sep/2021:16:03:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 5959 "-" "Mozilla 8.1"
192.168.8.241 - [24/Sep/2021:16:09:43 -0200] "GET /heartbeat.php HTTP/1.1" 200 3093 "-" "Mozilla 8.1"
192.168.10.133 - [24/Sep/2021:16:16:23 -0200] "GET /heartbeat.php HTTP/1.1" 200 352 "-" "Mozilla 8.1"
192.168.7.203 - [24/Sep/2021:16:23:03 -0200] "GET /heartbeat.php HTTP/1.1" 200 10641 "-" "Mozilla 8.1"
```





Probability distributions

Three types of random variables can be used:

- Discrete
- Continuous
- **Mixed** - continuous distributions that also have discrete outcomes.
Although useful in Performance Evaluation, we will not consider them for time constraints.



Probability distributions

Discrete random variables associate a set of probabilities to a discrete number of possible outcomes.

$$\Omega = \{red, green, blue\}$$

$$p(red) = 0.213$$

$$p(green) = 0.715$$

$$p(blue) = 0.072$$



Probability distributions

If the outcomes are numbers, the *Cumulative Distribution Function* can be defined: the probability $F(x)$ of getting an outcome less or equal to x .

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$p(1) = 0.15 \quad p(2) = 0.15$$

$$p(3) = 0.25 \quad p(4) = 0.15$$

$$p(5) = 0.15 \quad p(6) = 0.15$$

$$F(1) = 0.15 \quad F(2) = 0.30$$

$$F(3) = 0.55 \quad F(4) = 0.70$$

$$F(5) = 0.85 \quad F(6) = 1.00$$



In performance evaluation they are used to guide *random choices*.





Probability distributions

Continuous random variables generate outcomes on a continuous space.

$$\Omega = [0, \infty)$$

The probability of having an exact value is always zero.

$$p(X = x) = 0$$

However, we can compute the probability that an outcome is included in a range.

$$p(x_A \leq X \leq x_B) \geq 0$$



Probability distributions

They are characterized by the probability density function $f(x)$ and the cumulative distribution function $F(x)$.

$$p(x_A \leq X \leq x_B) = \int_{x_A}^{x_B} f(x) dx = F(x_B) - F(x_A)$$

The PDF should be such that its integral over the distribution support (generally, from 0 to ∞), is equal to one.

$$F(x) = \int_{-\infty}^x f(y) dy \quad \int_{-\infty}^{+\infty} f(y) dy = 1$$
$$f(x) = \frac{dF(x)}{dx}$$

Probability distributions

In performance evaluation they are usually adopted to characterize the inter-arrival times and the service times of the components of a model. For this reason they are generally defined only for positive values of the random variable.

$$F(x) = \int_0^x f(y)dy \quad \int_0^{+\infty} f(y)dy = 1$$





For a probability distribution, the *Expected value*, according to a function $g(x)$, is defined as:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

When $g(x)=x^n$, the corresponding expected value is called the n^{th} *moment* of the distribution.

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x)dx$$



Let us call $\mu = E[X]$ the first moment of the distribution. The n^{th} *central moment* of the distribution is defined as:

$$E[(X - \mu)^n]$$

Let us also call $\sigma = \sqrt{E[(X - \mu)^2]}$. The n^{th} *standardized moment* of the distribution is defined as:

$$E\left[\left(\frac{X - \mu}{\sigma}\right)^n\right]$$

Central moments are meaningful for $n > 1$, and are *insensitive to the position of the distribution*.

Standardized moments are meaningful for $n > 2$, and are *insensitive to both the scale and the position of the distribution*.



If we consider independent random variables, we can prove a lot of properties of the corresponding expected values:

$$E[c] = c$$

Constant

$$E[X + Y] = E[X] + E[Y]$$

Sum

$$E[aX] = aE[X]$$

Product with a constant

$$Y = \begin{cases} X_1 & p_1 \\ \dots & \dots \\ X_n & p_n \end{cases} \Rightarrow E[Y] = \sum_{i=1}^n p_i E[X_i]$$

Choice



The first moment is known as the *mean*, and corresponds to the average value of the distribution.

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

The second central moment is known as the *variance*.

$$Var[X] = E[(X - \mu)^2]$$

The variance can be computed from the second moment and from the mean with a simple expression:

$$\begin{aligned} Var[X] &= E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = \\ &= E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2 \end{aligned}$$



The square root of the variance is known as the *standard deviation*: its feature is that it uses the same units as the mean.

$$\sigma^2 = \text{Var}[X]$$

$$\sigma = \sqrt{\text{Var}[X]}$$

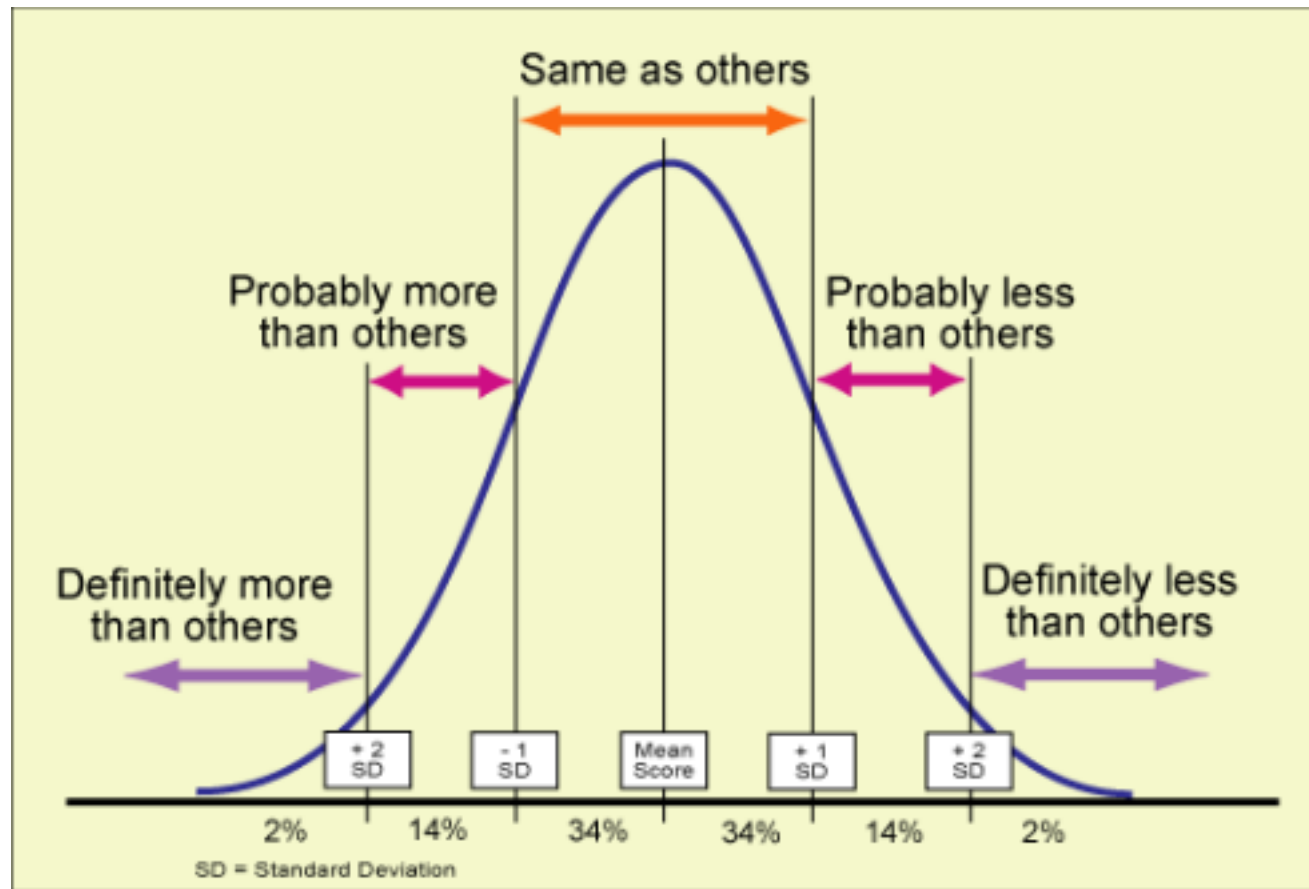
The ratio between the standard deviation and the mean is called the *coefficient of variation* (c_v).

$$c_v = \frac{\sigma}{\mu}$$

It tells how different the variance is with respect to the mean: we will return on the importance of this value later.



Variance, standard deviation and coefficient of variation express the same concept: the way in which the outcomes spread from the average, which represents the *variability* of the values.

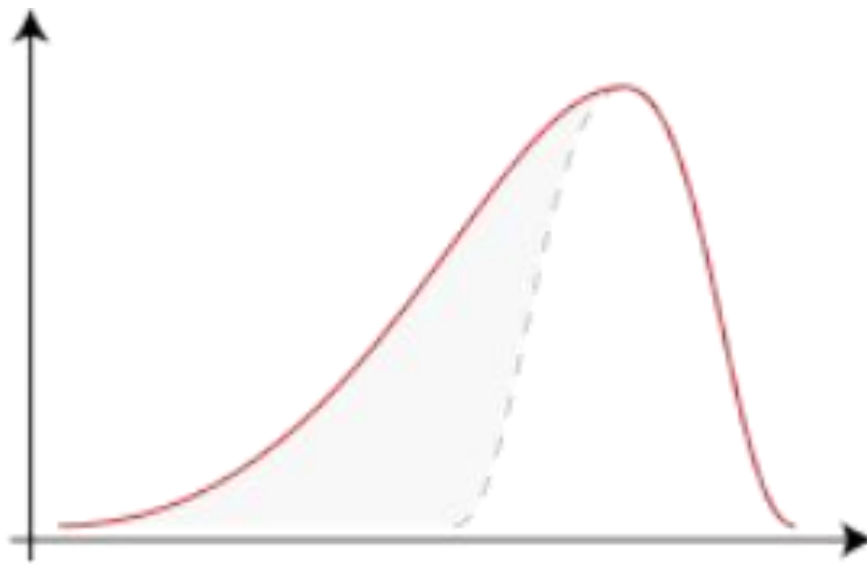




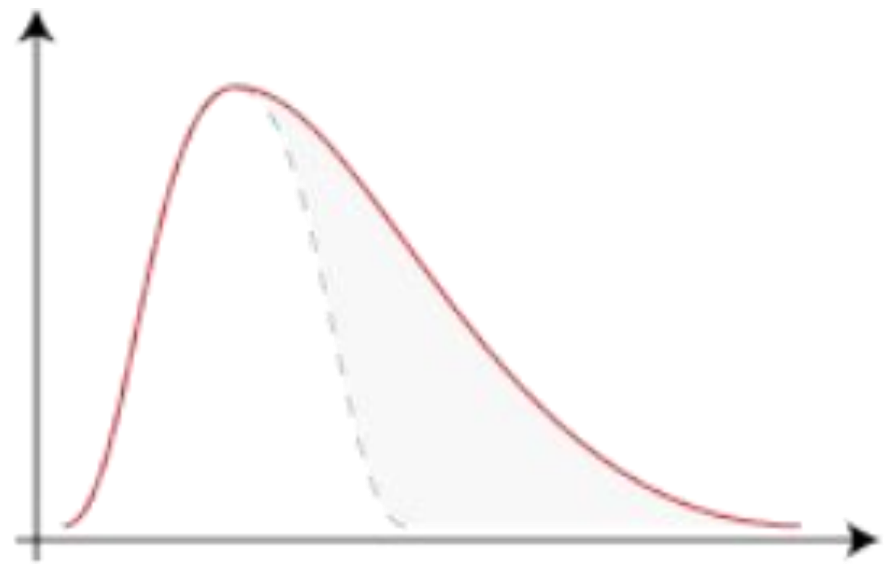
The third standardized moment is called the *skewness*.

$$\gamma = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{E[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

Skewness represents whether the distribution is symmetric, or if it has more probability mass to the left or to the right of its mean.



Negative Skew



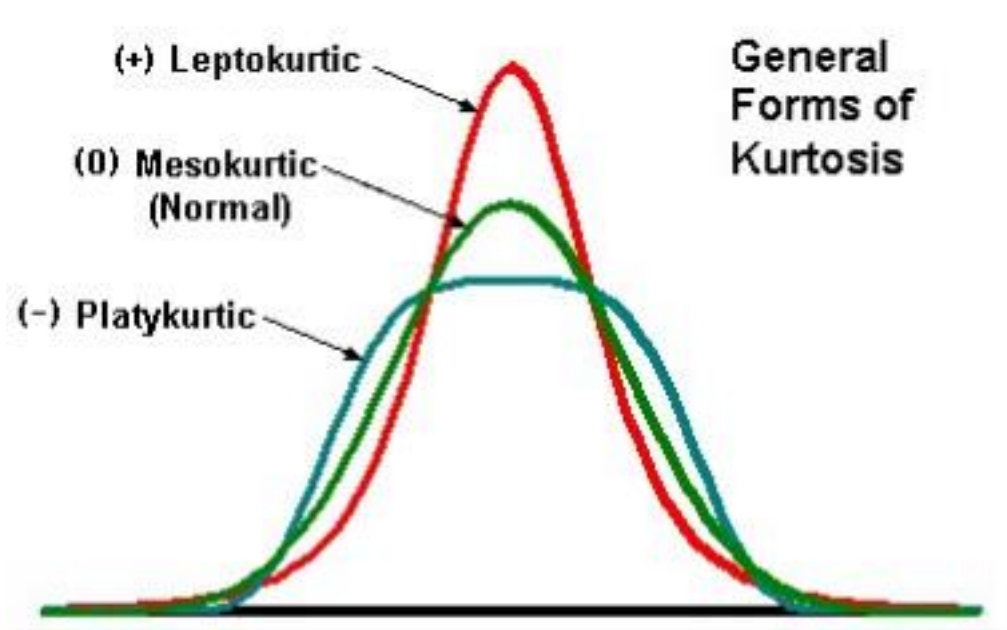
Positive Skew



The fourth standardized moment (minus 3) determines a characteristic of the distribution *called the excess Kurtosis*. The -3 is introduced to have the Kurtosis of the Standard Normal Distribution equal to 0.

$$\beta = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3 = \frac{E[(X - \mu)^4]}{\sigma^4} - 3$$

It describes the shape of the peak for the distribution.

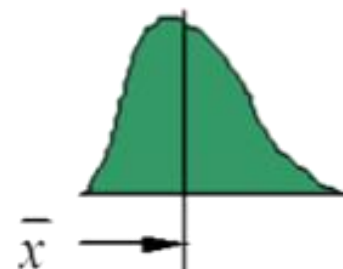




To summarize, the first four moments represent:

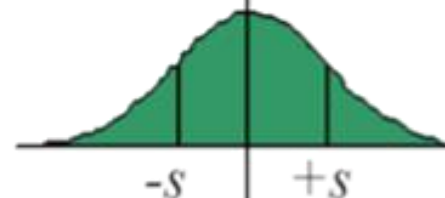
First Moment:

mean - measure of location



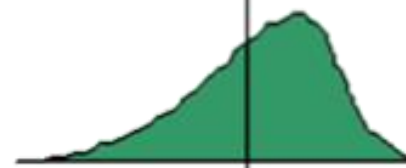
Second Moment:

Standard deviation - measure of spread



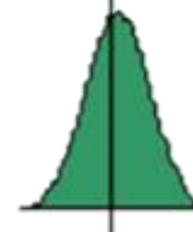
Third Moment:

skewness - measure of symmetry



Fourth Moment:

kurtosis - measure of peakedness





Moments are important in performance evaluation for several reasons:

- There are many results that are sensitive only to the first or one or two moments of the distributions.
- Beside being computed analytically on distributions, they can be easily derived from measures, log files and data sets.
- Moments of the collected data can guide the modeler to choose and parametrize the most appropriate distributions.



For most of the more popular probability distributions, moments can be computed analytically.

Luckily, many books and web resources that deals with probability distributions (including *Wikipedia*), present (when available) the analytical formula for the moments.

Exponential distribution - Wiki x

en.wikipedia.org/wiki/Exponential_distribution

Properties [\[edit \]](#)

Mean, variance, moments and median [\[edit \]](#)

The mean or **expected value** of an exponentially distributed random variable X with rate parameter λ is given by

$$E[X] = \frac{1}{\lambda}.$$

In light of the examples given [below](#), this makes sense: if you receive phone calls at an average rate of 2 per hour, then you can expect to wait half an hour for every call.

The **variance** of X is given by

$$\text{Var}[X] = \frac{1}{\lambda^2},$$

so the **standard deviation** is equal to the mean.

The **moments** of X , for $n \in \mathbb{N}$ are given by

$$E[X^n] = \frac{n!}{\lambda^n}.$$

The **central moments** of X , for $n \in \mathbb{N}$ are given by

$$\mu_n = \frac{!n}{\lambda^n} = \frac{n!}{\lambda^n} \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

where $!n$ is the **subfactorial** of n

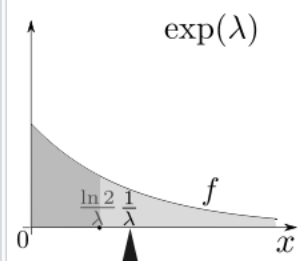
The **median** of X is given by

$$m[X] = \frac{\ln(2)}{\lambda} < E[X],$$

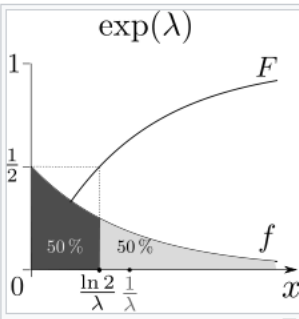
where \ln refers to the **natural logarithm**. Thus the **absolute difference** between the mean and median is

$$|E[X] - m[X]| = \frac{1 - \ln(2)}{\lambda} < \frac{1}{\lambda} = \sigma[X],$$

in accordance with the **median-mean inequality**.



The mean is the probability mass centre, that is the **first moment**.



The median is the **preimage** $F^{-1}(1/2)$.



Moments in finite discrete distributions

For distributions on discrete support, moments, mean, variance and other properties can be computed as finite sums.

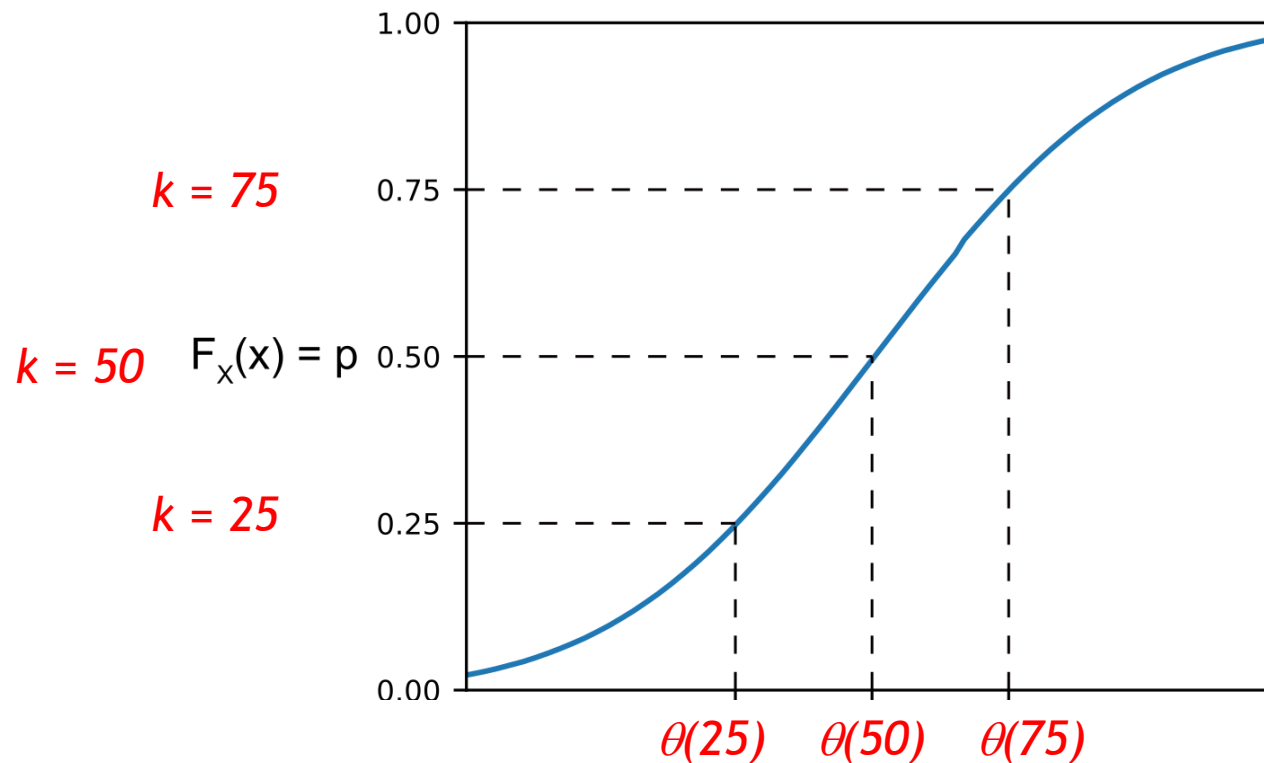
$$\mu = E[X] = \sum_i x_i p(x_i) \quad E[X^k] = \sum_i x_i^k p(x_i)$$

$$Var[X] = E[X^2] - \mu^2 = \sum_i x_i^2 p(x_i) - \mu^2$$



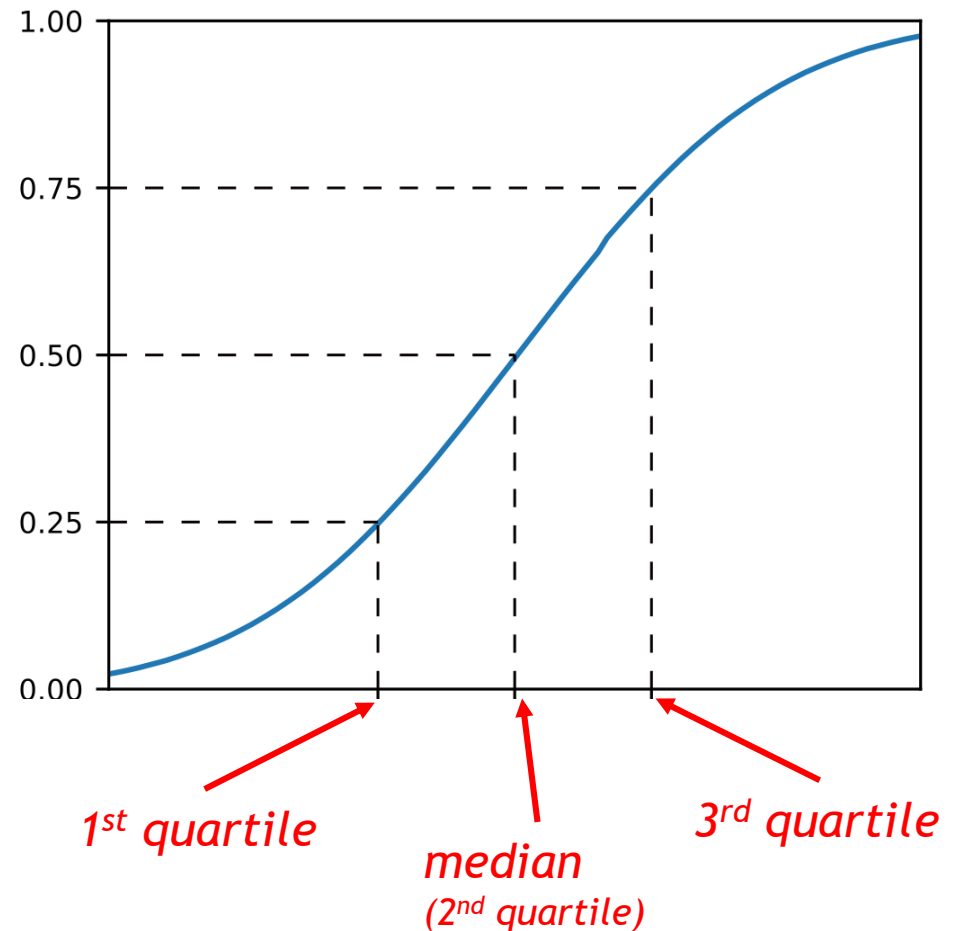
The k^{th} percentile of a distribution is the value of the random variable, for which the CDF equals $k/100$:

$$\theta(k) = F^{-1}(k/100)$$





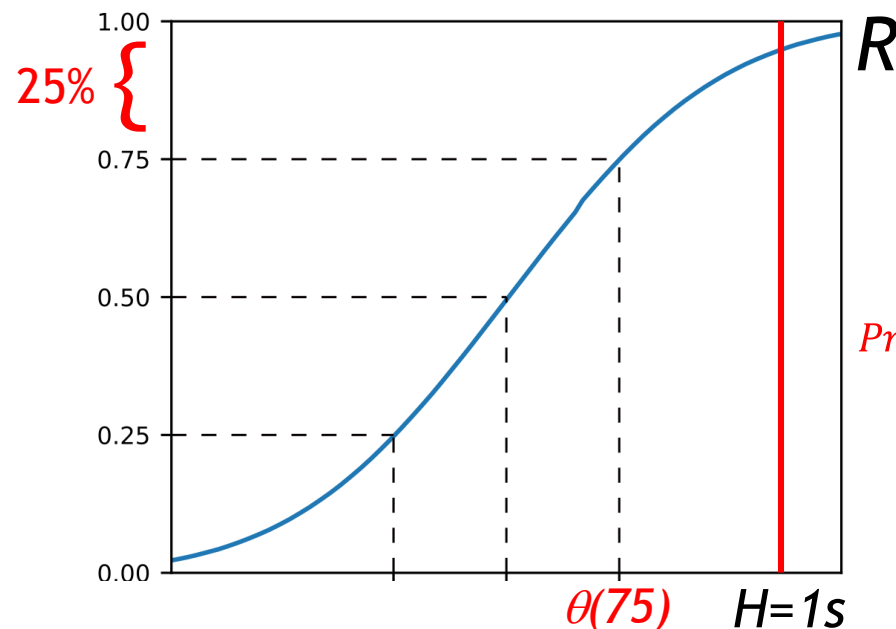
The 25th, 50th and 75th percentiles have special names: they are respectively called the *1st quartile*, the *median* (or *2nd quartile*), and the *3rd quartile*.





In performance evaluation percentiles are used to assess properties: for example, they can be used to ensure that the response time R of a station is greater than a given threshold H only for a limited percentage k of jobs.

$$\text{Prob}(R \geq H) \leq \frac{k}{100} \Leftrightarrow \theta(100 - k) \leq H$$



$$\text{Prob}(R \geq 1s) \leq 0.25 \Leftrightarrow \theta(75) \leq 1s$$



Distributions from traces

As introduced, probability distributions can help characterizing the workload trace.

Let us see which relations we can find between traces and probability distributions.



CDF from samples of a distribution

Estimating the PDF of a distribution from a set of samples is not an easy task.

Approximating the CDF is instead much simpler.

Let us assume that we have N samples, $x_1 \dots x_N$ from an unknown distribution X .

The screenshot shows the Microsoft Excel interface with the 'Formulas' tab selected. The formula bar displays the value 0,229267082953235. The worksheet contains a table with columns A, B, and C. Column A is empty, column B contains cumulative values, and column C is empty. The values in column B are: 0,2292671, 0,1821856, 0,1962806, 0,2733049, 0,1098125, 0,1310496, 0,1171538, 0,08043, 0,3319404, 0,1568503, 0,1035784, 1,0186532, 0,2255764, 0,1023485, 0,0659296, 0,2592162, 0,6156334, 0,4451787, 0,2264959, 0,262116.

	A	B	C
1			
2		0,2292671	
3		0,1821856	
4		0,1962806	
5		0,2733049	
6		0,1098125	
7		0,1310496	
8		0,1171538	
9		0,08043	
10		0,3319404	
11		0,1568503	
12		0,1035784	
13		1,0186532	
14		0,2255764	
15		0,1023485	
16		0,0659296	
17		0,2592162	
18		0,6156334	
19		0,4451787	
20		0,2264959	
21		0,262116	
22			
23			



CDF from samples of a distribution

The CDF of X can be approximated by first sorting the samples: let us call $y_1 \dots y_N$ the sorted version of samples $x_1 \dots x_N$.

$$|y_1, \dots, y_N| = \text{sort}(x_1, \dots, x_N)$$
$$y_1 \leq \dots \leq y_N$$

	A	B	C	D	E
1					
2		0,2292671		0,0659296	
3		0,1821856		0,08043	
4		0,1962806		0,1023485	
5		0,2733049		0,1035784	
6		0,1098125		0,1098125	
7		0,1310496		0,1171538	
8		0,1171538		0,1310496	
9		0,08043		0,1568503	
10		0,3319404		0,1821856	
11		0,1568503		0,1962806	
12		0,1035784		0,2255764	
13		1,0186532		0,2264959	
14		0,2255764		0,2292671	
15		0,1023485		0,2592162	
16		0,0659296		0,262116	
17		0,2592162		0,2733049	
18		0,6156334		0,3319404	
19		0,4451787		0,4451787	
20		0,2264959		0,6156334	
21		0,262116		1,0186532	
22					
23					



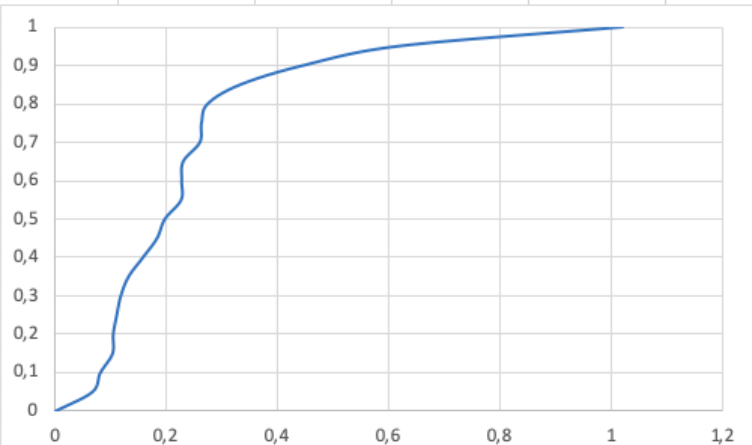
CDF from samples of a distribution

Then, the approximated CDF of X can be defined as:

$$F_X(x) = \frac{i}{N} \quad \forall y_i \leq x < y_{i+1} \quad (\text{with } y_0 = -\infty, y_{N+1} = +\infty)$$

Using the indicator function $I(\phi)$, we have:
$$F_X(x) = \frac{1}{N} \sum_{i=1}^N I(y_i \leq x)$$

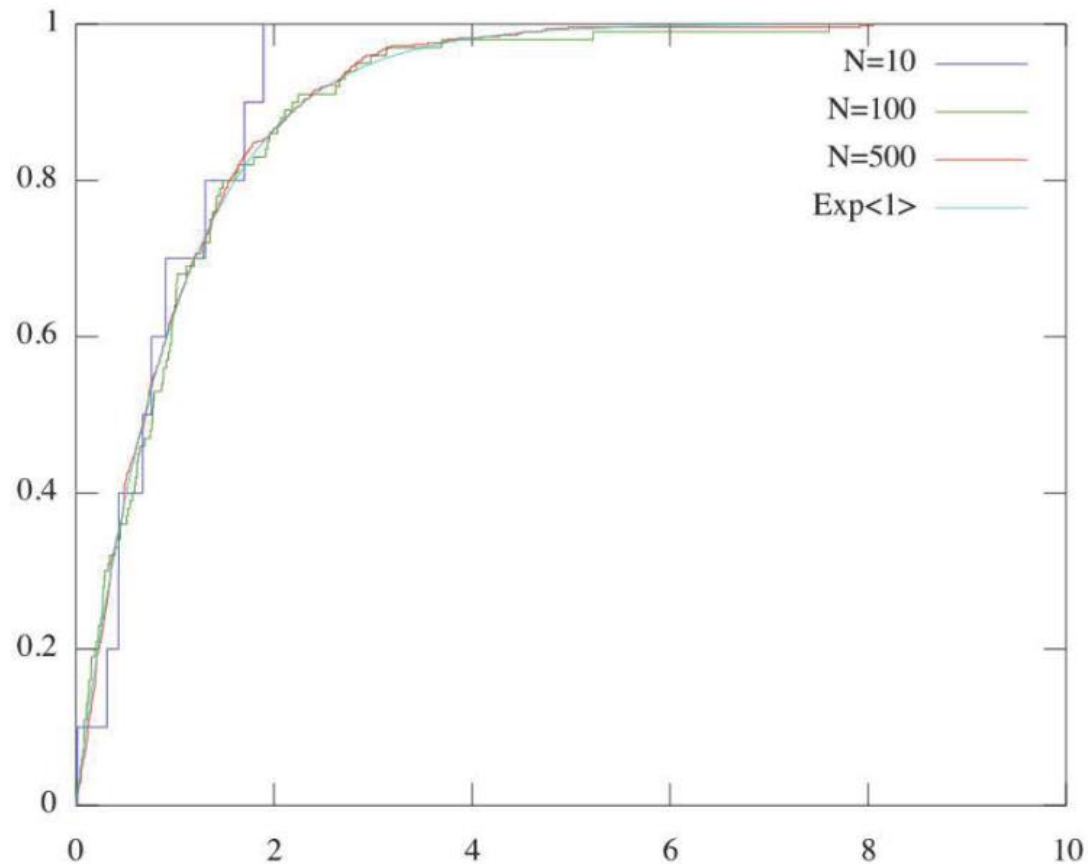
0,2292671	0,0659296	0,05
0,1821856	0,08043	0,1
0,1962806	0,1023485	0,15
0,2733049	0,1035784	0,2
0,1098125	0,1098125	0,25
0,1310496	0,1171538	0,3
0,1171538	0,1310496	0,35
0,08043	0,1568503	0,4
0,3319404	0,1821856	0,45
0,1568503	0,1962806	0,5
0,1035784	0,2255764	0,55
1,0186532	0,2264959	0,6
0,2255764	0,2292671	0,65
0,1023485	0,2592162	0,7
0,0659296	0,262116	0,75
0,2592162	0,2733049	0,8
0,6156334	0,3319404	0,85
0,4451787	0,4451787	0,9
0,2264959	0,6156334	0,95
0,262116	1,0186532	1





CDF from samples of a distribution

For example, for the exponential distribution with 10, 100 and 500 samples we have:





Moments from samples

Assuming that each sample is equally probable, moments can be approximated with discrete sums.

$$\mu = E[X] = \frac{1}{N} \sum_{i=1}^N x_i \quad E[X^k] = \frac{1}{N} \sum_{i=1}^N x_i^k$$

Please note that statisticians have derived more accurate version of many of these concepts starting from samples. For example, the variance is better defined as:

$$Var[X] = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2$$

In Performance Evaluation, since we generally have a very large data set, the difference between the two versions is minimal. We can then use the simplified versions presented here, without affecting the results.



Percentiles from traces

Several approaches exist to determine percentiles from samples.
For an exhaustive list, please refer to:

<https://axibase.com/use-cases/workshop/percentiles.html#discontinuous-sample>

One of the simplest to understand and to implement uses *linear interpolation*:

$$h = (N - 1) \frac{k}{100} + 1$$
$$\theta(k) = \begin{cases} y_N & h = N \\ y_{\lfloor h \rfloor} + (h - \lfloor h \rfloor) \cdot (y_{\lfloor h \rfloor + 1} - y_{\lfloor h \rfloor}) & h < N \end{cases}$$

Please note that we used y_h , to denote that samples need to be sorted for this method to be used, and that the indices of the samples starts from 1.



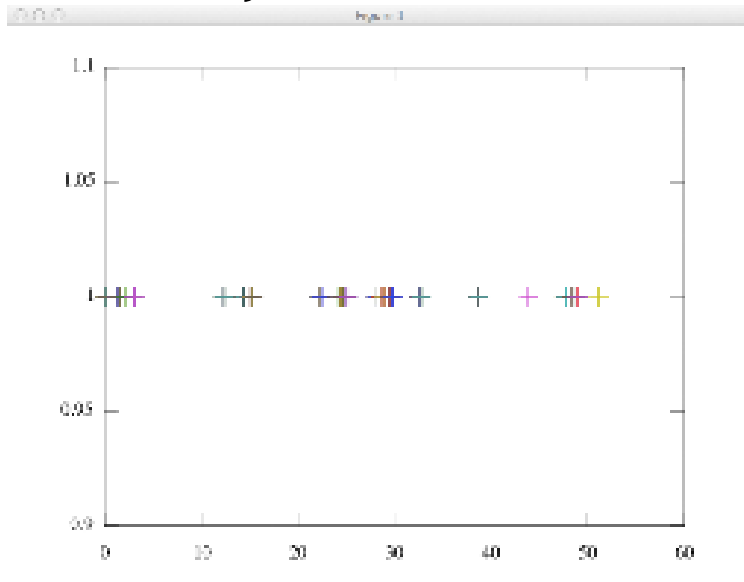
The correlation of a trace can be identified with the *cross-covariance* among successive samples of the considered distribution, comparing each one with the value m jobs later.

If we call X_n the n -th instance of a given distribution, the *lag- m cross-covariance* $\sigma(m)$ is defined as:

$$\sigma(m) = E[(X_n - \mu_X)(X_{n+m} - \mu_X)]$$

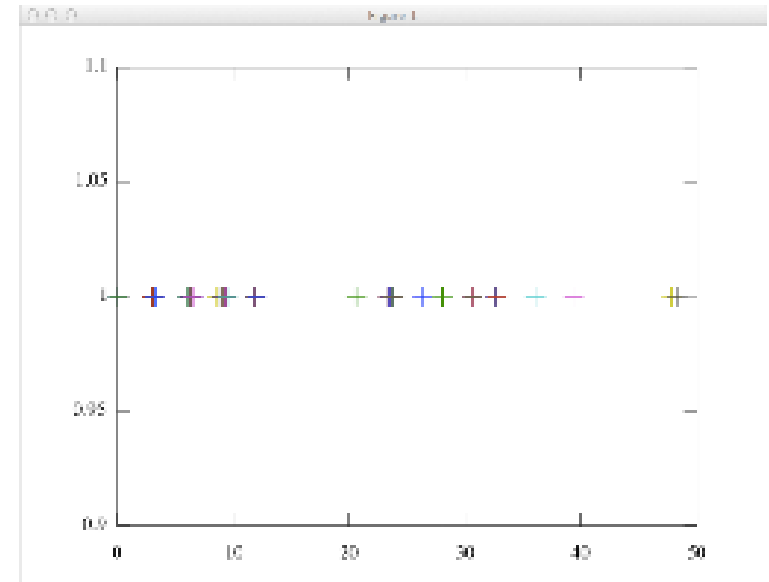


If there is no correlation, the cross covariance tends to zero. If it is very different from zero, then samples are correlated.



$$\sigma(1) \sim 0$$

Hyper(0.025, 2.4, p=0.8)
 $R_3(0.35) = 4.83 \text{ s}$



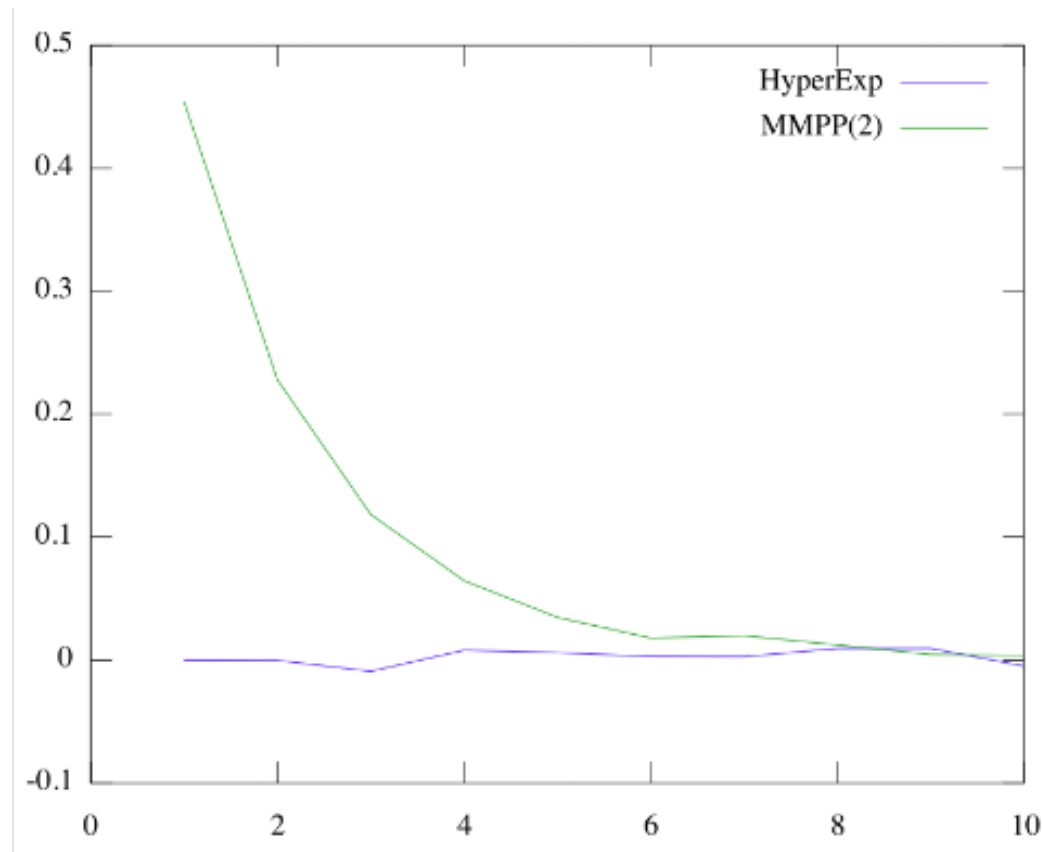
$$\sigma(1) \sim 0.5$$

MMPP(2)
 $R_4(0.35) = 9.19 \text{ s}$



Parameter m represents the "distance" between two samples. In general the cross-covariance tends to zero as m increases.

The slower it tends to zero, the more correlated the samples are.





Cross-covariance

Cross-covariance to check correlation of a data-set, can be directly computed, since such measure is already defined on samples of a distribution:

$$\sigma(m) \cong \frac{1}{N-m} \sum_{n=1}^{N-m} (x_n - \mu_X)(x_{n+m} - \mu_X)$$

Please note that here we used x_n to denote that samples *must NOT* be sorted for this method to work.



Cross-covariance

Cross-covariance has the same scale problems as the variance.

The normalized version of the Cross-Covariance (i.e. the equivalent of the coefficient of variation), is called the *Pearson Correlation Coefficient*:

$$\rho(m) = \frac{E[(X_n - \mu_X)(X_{n+m} - \mu_X)]}{E[(X_n - \mu_X)^2]}$$

$$\rho(m) \cong \frac{\frac{1}{N-m} \sum_{n=1}^{N-m} (x_n - \mu_X)(x_{n+m} - \mu_X)}{\frac{1}{N} \sum_{n=1}^N (x_n - \mu_X)^2}$$



Analysis of Motivating Example

Using the formulas just seen, and a mathematical package, we can compute average, standard deviation and cross-covariance to determine intensity, variability and correlation of the considered trace.

The image shows the MATLAB R2017b - academic use interface. The top menu bar includes HOME, PLOTS, and APPS. Below the menu bar is a toolbar with icons for New Script, New Live Script, Find Files, and Compare. The main workspace area is divided into three panes: Current Folder, Command Window, and Workspace.

Current Folder: The file browser shows a directory structure with files like Workload types.docx, MMPP2.global.csv, HyperExponential.glob..., http log simulator, Exponential.global.csv, Deterministic.global.csv, BaseL02-2022.jsimg, Assignment and Datas..., WK - Assignment and ..., and Assignment and Dataset.

Command Window: The Command Window displays the following MATLAB code and its output:

```
>> A=csvread("MMPP2.global.csv",1,1);
>> B=diff(A(:,1));
>> N=size(B,1);
>> MeanB = sum(B)/N

MeanB =

    1.2723

>> StdVarB = sqrt(sum((B - MeanB).^2)/N)

StdVarB =

    2.1577

>> Lag1B = sum((B(1:N-1,1)-MeanB).*(B(2:N)-MeanB))/(N-1)

Lag1B =

    2.0236

fx >>
```

Workspace: The Workspace pane shows the following variables and their sizes:

Variable	Size
A	1x1
Aavg	2
Acv	2
Afs	1x1
Aprocess	0
Arho	0