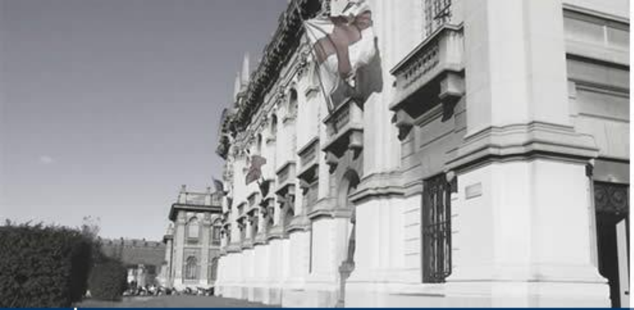


Performance Evaluation and Applications

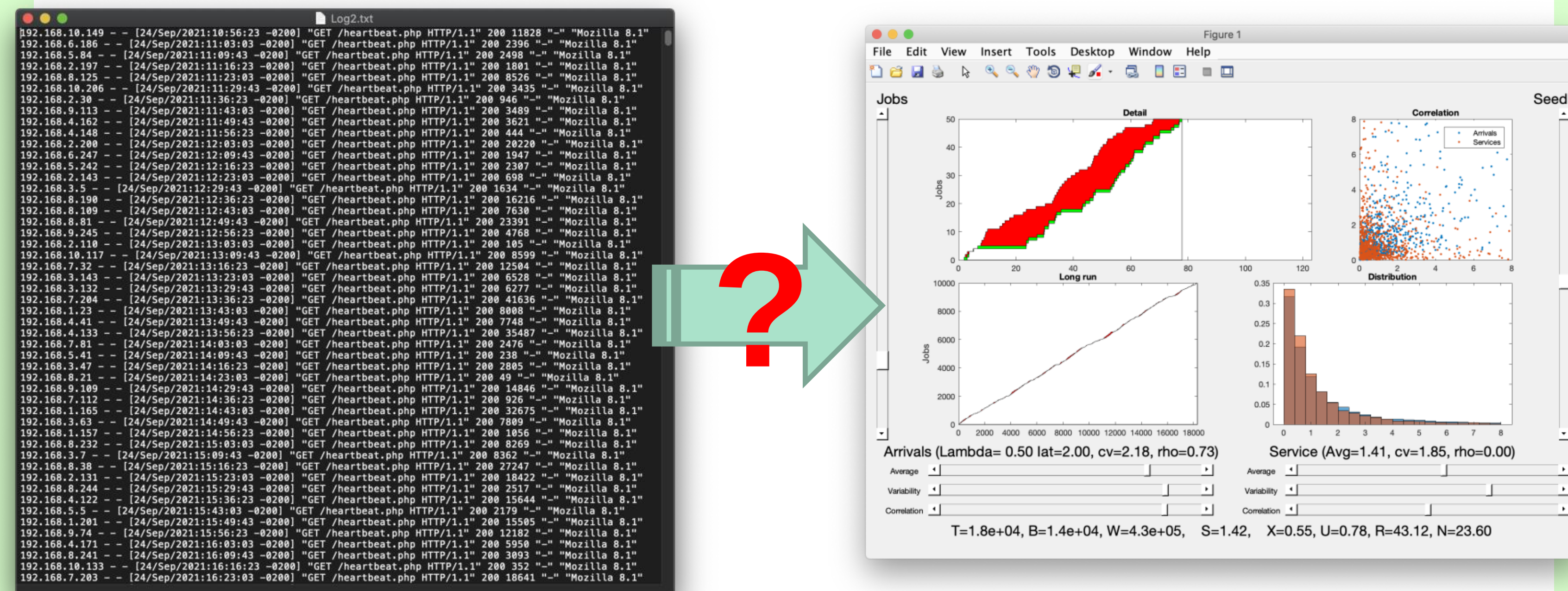
POLITECNICO DI MILANO

Modeling Workloads

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?



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*.

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

$$p(1) = 0.15 \quad p(2) = 0.15$$
$$p(3) = 0.25 \quad p(4) = 0.15$$
$$p(5) = 0.15 \quad p(6) = 0.15$$

$$F(1) = 0.15 \quad F(2) = 0.30$$
$$F(3) = 0.55 \quad F(4) = 0.70$$
$$F(5) = 0.85 \quad F(6) = 1.00$$

Probability distributions

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.

$$F(x) = \int_{-\infty}^{x} f(y)dy \qquad \int_{-\infty}^{+\infty} f(y)dy = 1$$

$$f(x) = \frac{dF(x)}{dx}$$

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.

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

Moments

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.

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

Moments

Let us call μ = 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*.

Moments

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

Moments

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

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

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:

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

Moments

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

$$\sigma^2 = Var[X]$$

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

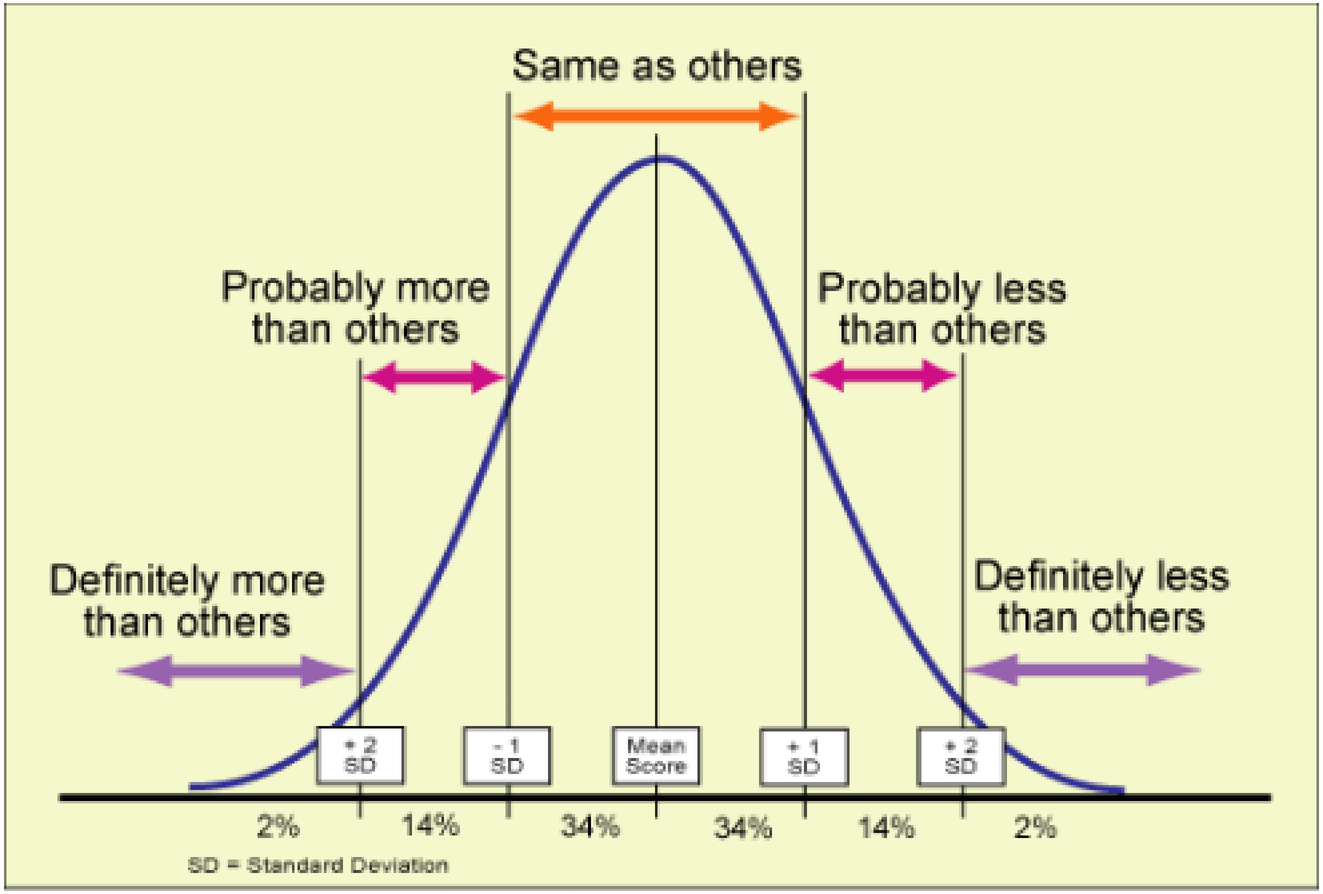
The ratio between the standard deviation and the mean is called the *coefficient of variation* (c_v).

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

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

Moments

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.

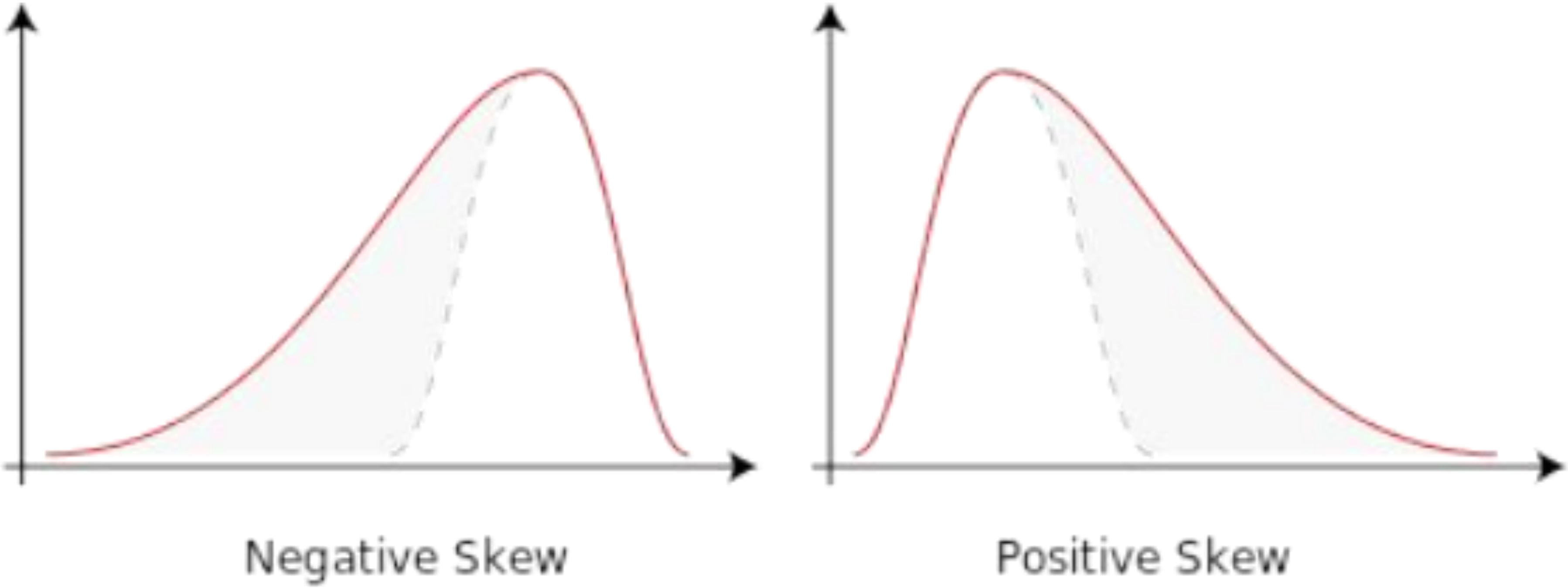


Moments

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.

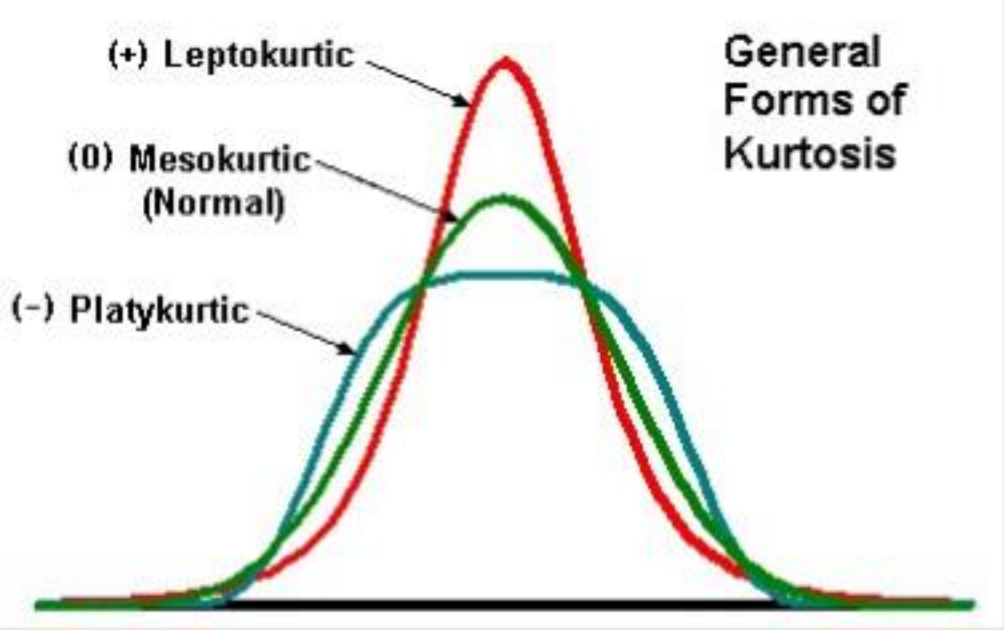


Moments

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.



Moments

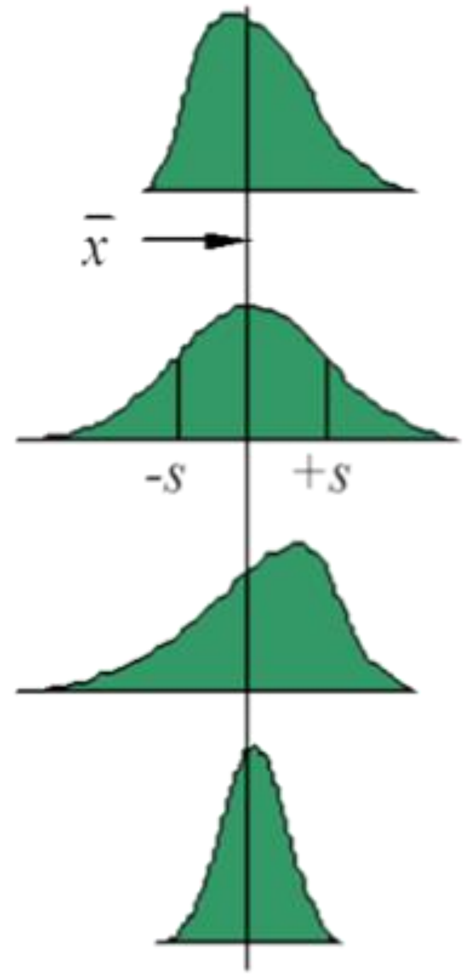
To summarize, the first four moments represent:

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

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.

Moments

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

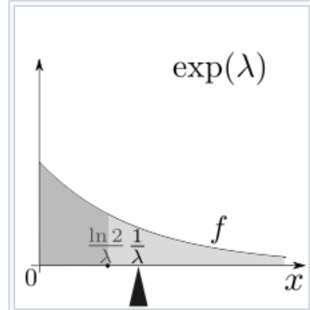
en.wikipedia.org/wiki/Exponential_distribution

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

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



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

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

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

so the standard deviation is equal to the mean.

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

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

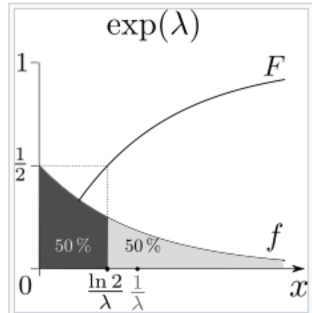
The central moments of X, for $n \in \mathbb{N}$ are given by

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

where $!n$ is the subfactorial of n

The median of X is given by

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



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

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

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

in accordance with the median-mean inequality.

Moments in finite discrete distributions

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

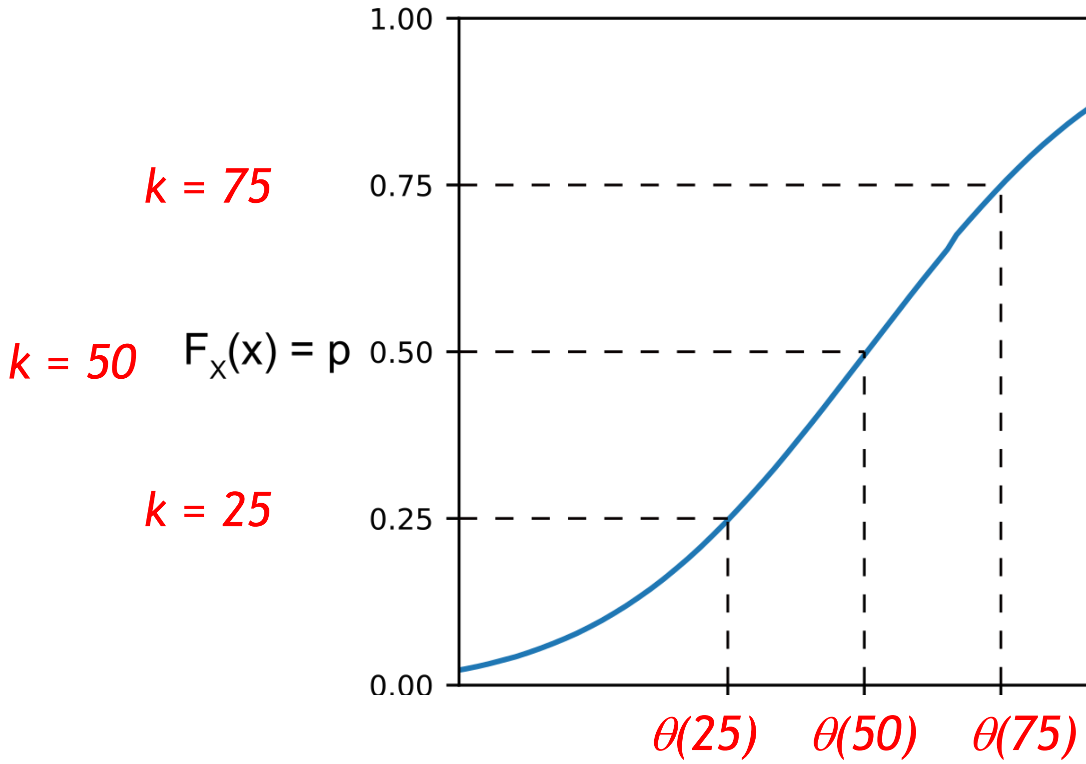
$$\mu = E[X] = \sum_i x_i p(x_i) \qquad E[X^k] = \sum_i {x_i}^k p(x_i)$$

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

Percentiles

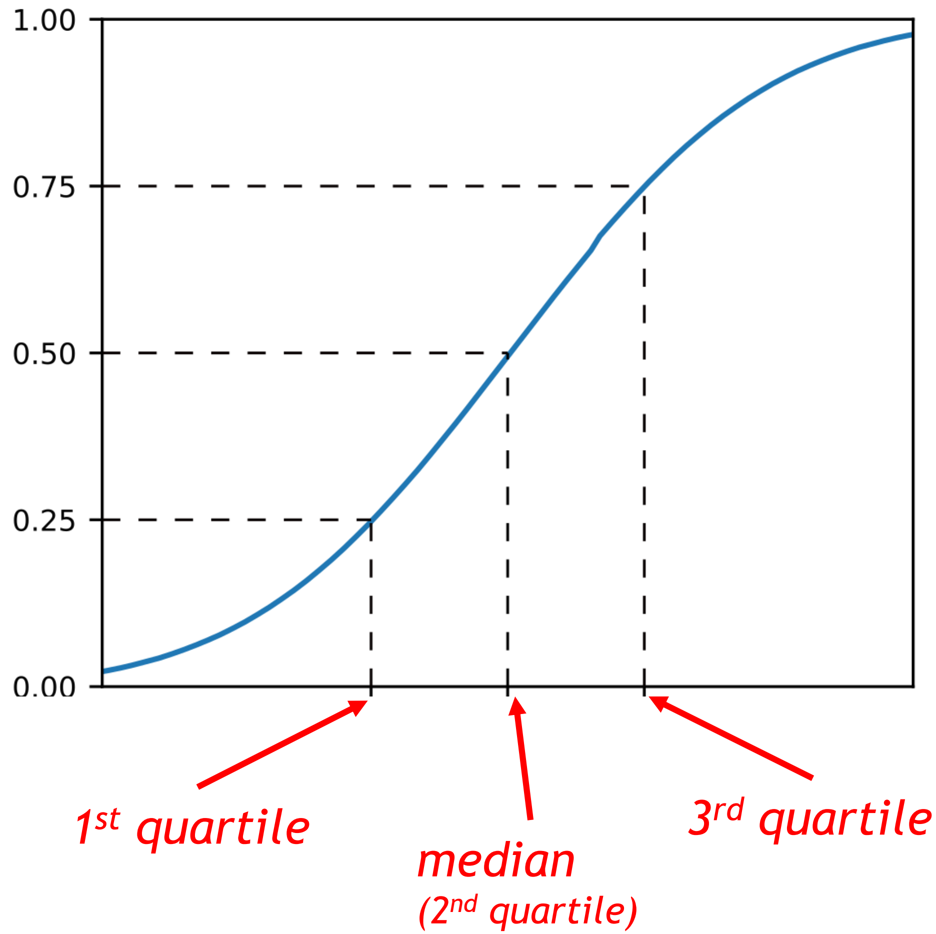
The k^{th} *percentile* of a distribution is the value of the random variable, for which the CDF equals *k/ 100*:

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



Percentiles

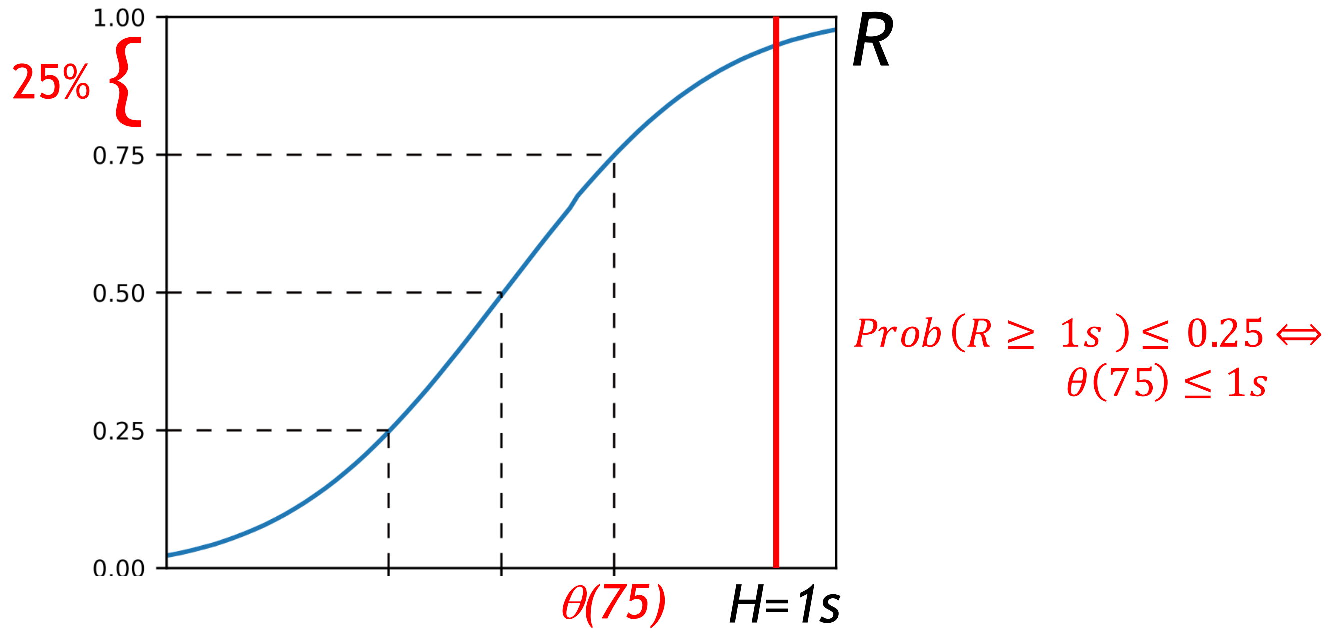
The 25th, 50th and 75th percentiles have special names: they are respectively called the *1st quartile*, the *median* (or 2nd *quartile*), and the *3rd quartile*.



Percentiles

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.

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



Distributions from traces

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

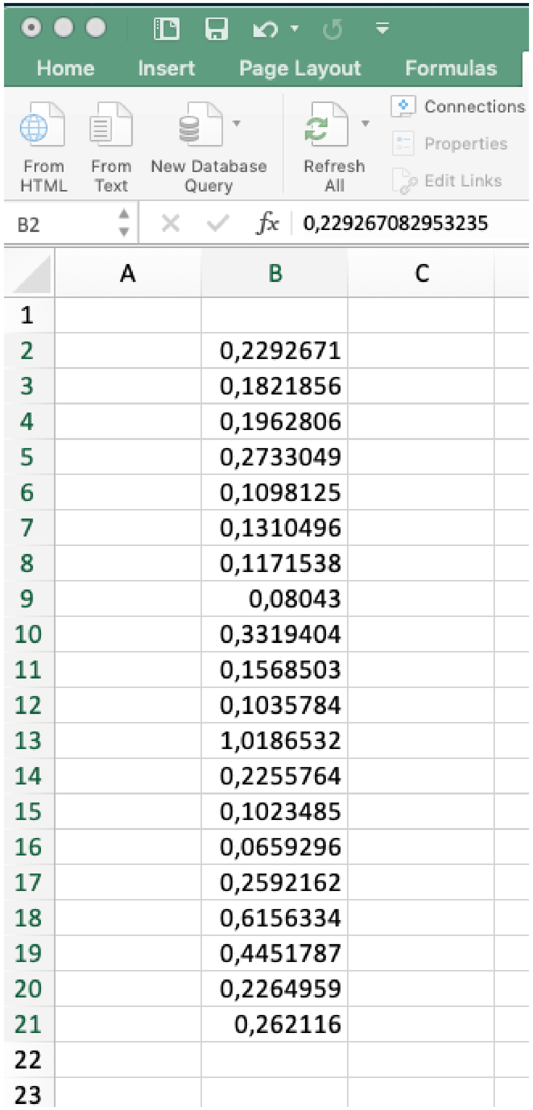
Let us see which relations we can find between traces and probability distributions.

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.

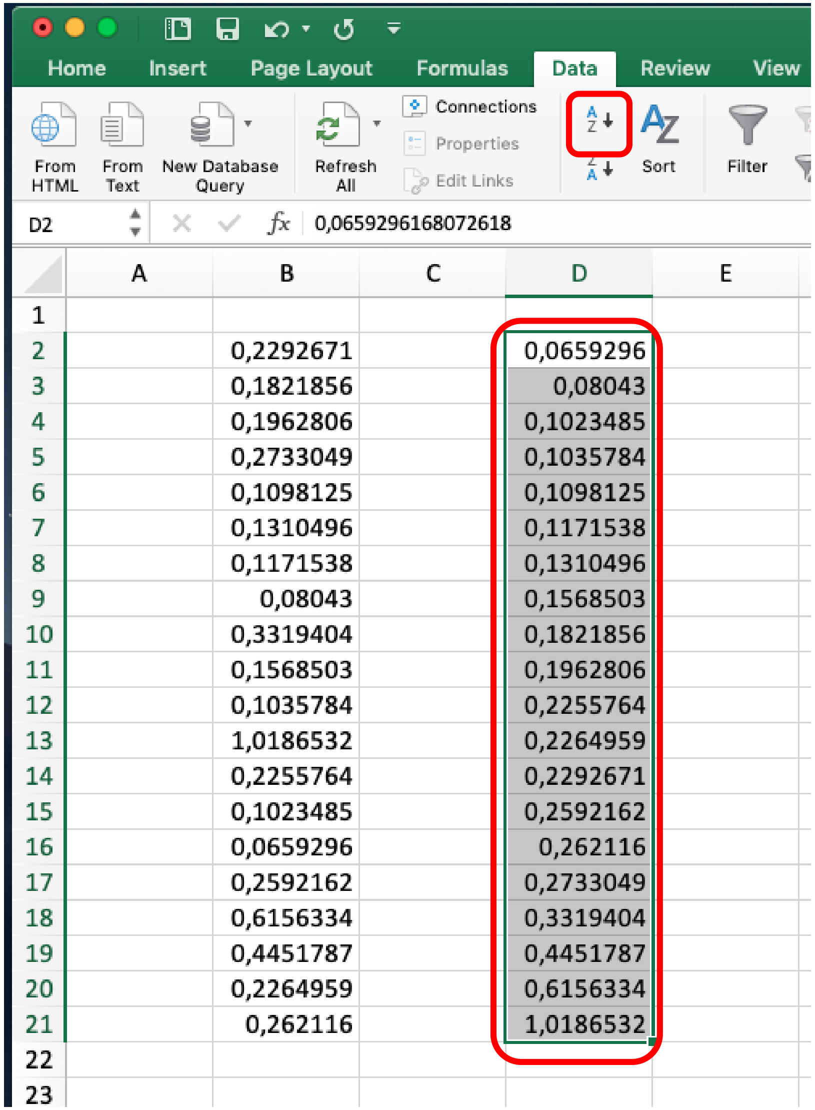
Let us assume that we have N samples, $x_1 \dots x_N$ from an unknown distribution X.



CDF from samples of a distribution

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

$$|y_1, \ldots, y_N| = sort(x_1, \ldots, x_N)$$
$$y_1 \leq \cdots \leq y_N$$



	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					

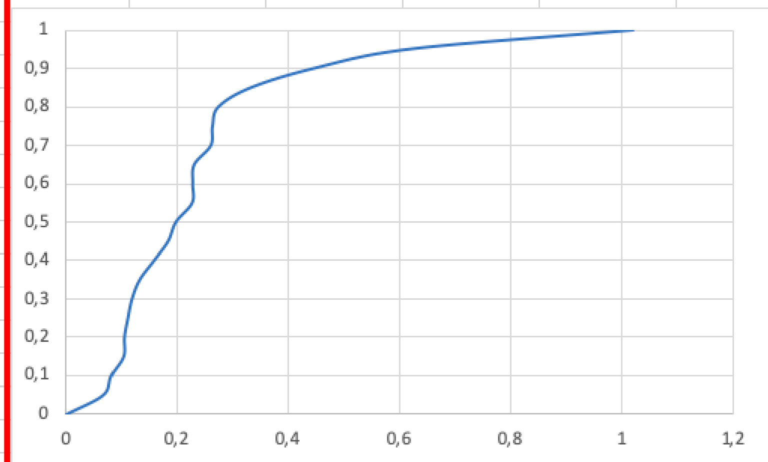
CDF from samples of a distribution

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

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

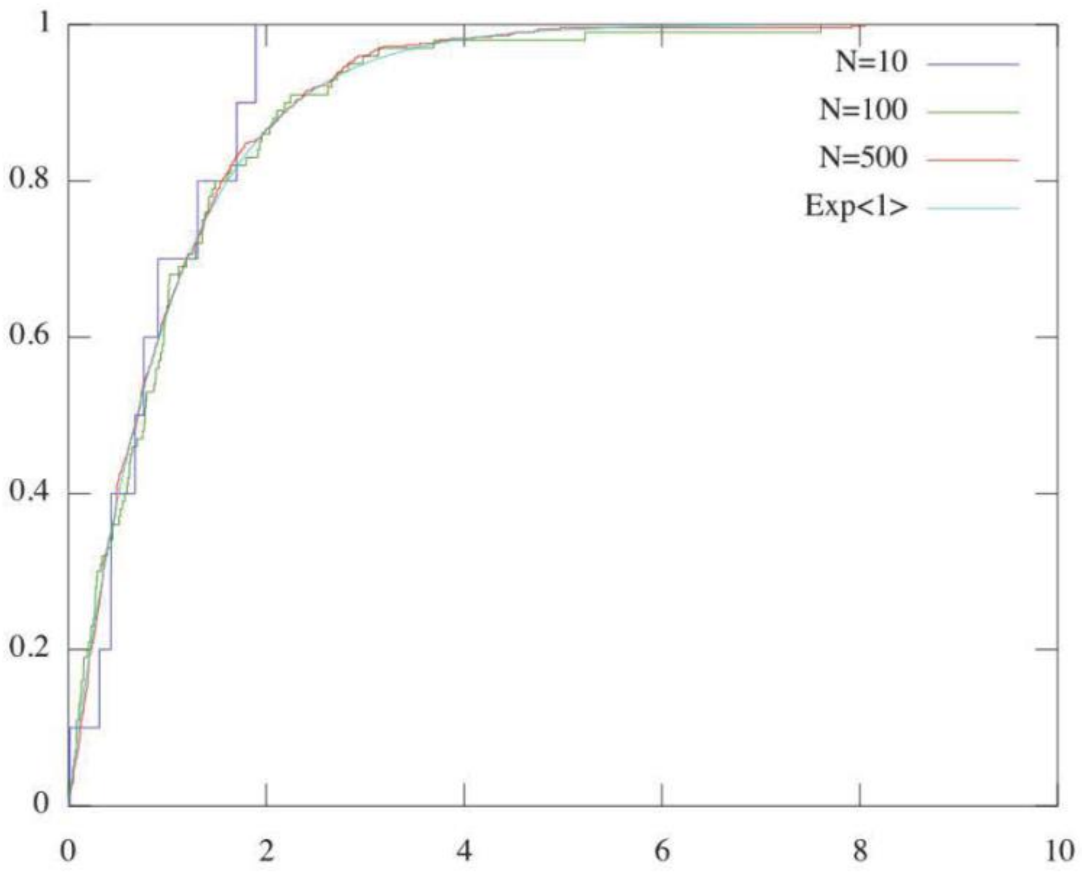
Using the indicator function *I(ϕ)*, we have: $$F_X(x) = \frac{1}{N}\sum_{i=1}^{N} I(y_i \leq x)$$

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.

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

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*:

$$h = (N-1)\frac{k}{100} + 1$$

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

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.

Correlation

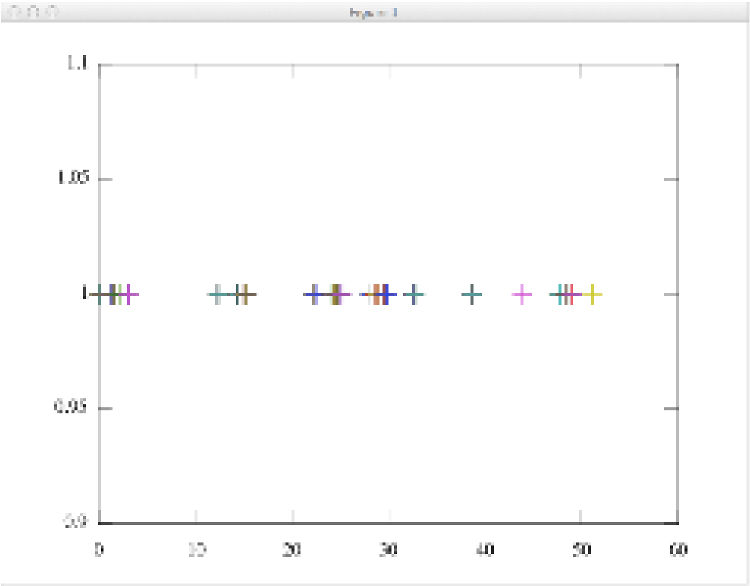
The correlation of a trace can be identified with the *cross-covariance* among successive samples of the considered distribution, comparing each one with the value *m* jobs later.

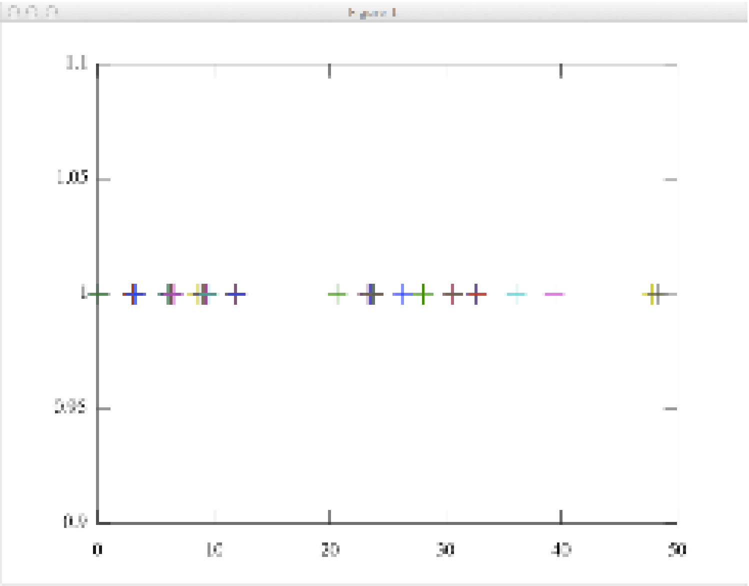
If we call X_n the n-th instance of a given distribution, the *lag-m cross-covariance* $\sigma(m)$ is defined as:

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

Correlation

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





$\sigma(1) \sim 0$

$\sigma(1) \sim 0.5$

Hyper(0.025,2.4,p=0.8)

$R_3(0.35) = 4.83\ s$

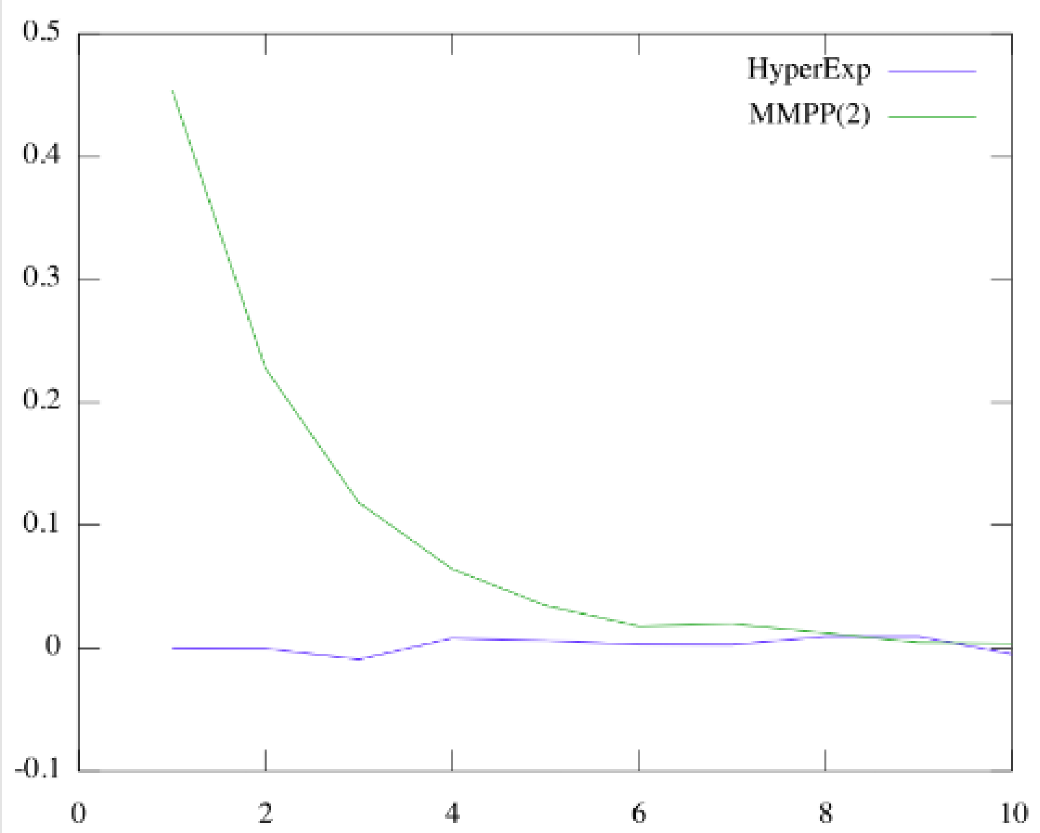
MMPP(2)

$R_4(0.35) = 9.19\ s$

Correlation

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

The slower it tends to 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.