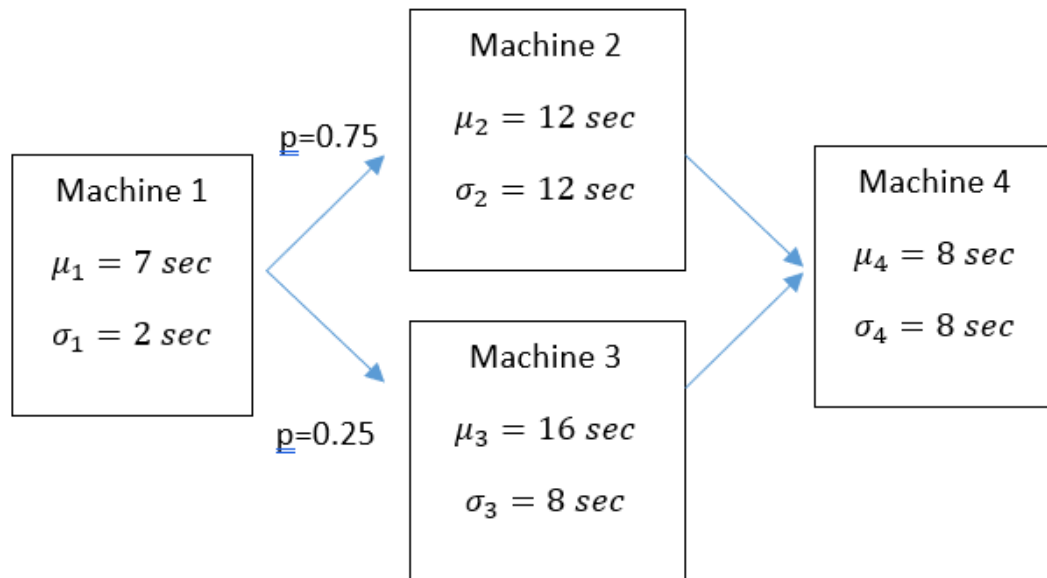


simulations, descriptive stats & estimation

This project illustrates the course with an industrial applications of simulations. You will analyse the performance of an assembly chain, analytically and with simulations. This is an introduction to supply chain management which is a field actively studied by engineers.

In a manufacture of mechanical devices for the automotive industry, the production of a gearbox is done by a line of 3 successive working stations. The production of a new gearbox starts only when the previous one is fully assembled. In what follows, the assembly time of one gearbox is denoted by T . Also, T_k is the assembly time at machine k . Finally $\mu_k = \mathbb{E}(T_k)$ and $\sigma_k^2 = \mathbb{V}(T_k)$. Details about the production line are provided in the graph below.



where $p = 0.75$ indicates for example that the output of machine 1 ends up in machine 2 with probability 0.75.

-
1. For each machine k , choose a (non-Gaussian) statistical distribution for the assembly time T_k and provide value(s) for its parameter(s).

Answer here

- b) Calculate the expected time for assembling one gearbox, $\mathbb{E}(T)$.

Answer here

1. Simulate 5000 durations T for assembling one single gearbox and report the mean (check that its is well close to the expectation of question 1), standard deviation, 5% and 95% percentile. Plot an histogram. Save the simulated durations in a file "sample.csv".

In [1]:

```
#Code here
```

1. a) Load simulated data from the file "sample.csv". Fit the exponential, gamma and chi-square distributions and estimate their parameters by log-likelihood maximization. Select the best non-Gaussian statistical distribution for T . Justify your choice.

Rem: if you use scipy, set the location parameter to zero with option 'floc=0'.

Answer here

In [2]:

```
#code here
```

- b) During the lectures, we have seen that the chi-square distribution has only one single parameter (the number of degrees of freedom). In Scipy, we find two other parameters location "loc" and scale "scale". Why?

Answer here

- c) Compare the empirical distribution with the distribution you chose at 3a). For this purpose, plot an histogram of durations and add the probability density function (pdf) that you have selected for modelling the random variable.
-

In [3]:

```
#code here
```

1. We want to fit a Weibull distribution to T . The density, $f(t)$, depends upon two parameters $\alpha > 0$ and $\lambda > 0$ and is given by:

$$f(t) = \lambda \alpha t^{\alpha-1} \exp(-\lambda t^\alpha)$$

- a) Write two functions: one for the density and one that computes the log-likelihood of a sample of durations for this distribution. Test this function on the simulated dataset and for $\alpha = 1$ and $\lambda = \frac{1}{28}$
-

In [4]:

```
#code here
```

- b) Maximize the log-likelihood to find estimates $\hat{\alpha}$ and $\hat{\lambda}$. Next compare the log-likelihood with those computed previously. Is this model better?

Hint: use the "scipy.optimize" package and the function "minimize". To avoid adding constraints, you can do a change of variables: $\alpha = \exp(a)$ and $\lambda = \exp(b)$. The function "minimize" may send back a warning (division by zero in log) but converges well to the solution.

In [5]:

```
#code here
```

c) The expectation and variance of the Weibull distribution are

$$\mathbb{E}(T) = \lambda^{-\frac{1}{\alpha}} \Gamma\left(\frac{1}{\alpha} + 1\right)$$

$$\mathbb{V}(T) = \lambda^{-\frac{2}{\alpha}} \left(\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma\left(\frac{1}{\alpha} + 1\right)^2 \right)$$

Estimate α and λ by the methods of moments. Next calculate the log-likelihood with these estimates and compare with these obtained by log-likelihood maximization. Is the result surprising ?

Hint: use the `scipy.optimize` package and the function "root". The gamma function is `scipy.special.gamma(z)`

In [6]:

```
#code here
```

1. Based on the content of the "Self-learning 2", Bootstrap 1000 times the sample of durations simulated in Question 2 (stored in the file "sample.csv") and compute, the mean, standard deviation and 95% and 90% confidence interval for parameter(s) estimate(s). Plot the empirical distribution of parameter(s).

In [7]:

```
#code here
```

1. Using simulated assembling times, determine the probability that assembling 10 gearboxes lasts less than 320 seconds $P(T \leq 320 \text{ sec})$.

In [8]:

```
#code here
```

1. Is there any alternative to numerical computing in question 6 for approximating $P(T \leq 320 \text{ sec})$? Please detail your answer and provide an estimate of this probability.

Answer here

In []: