# Applied Stochastic Processes
# Assignment I

## PART I - Theory

A high-tech data processing unit at a research lab is responsible for executing a computational task multiple times each day. However, the number of times $X$ that this task must be processed is **random** and varies from day to day. The **expected** number of executions $\mu$ is itself uncertain, as it depends on factors like incoming data volume and system workload. This expectation follows a **Gamma distribution** with parameters $\alpha$ (shape) and $\beta$ (rate).

The random variable $X$ is hence modelled according to the following hierarchical structure:

$$\mu \sim \Gamma(\alpha, \beta) \quad \text{and} \quad X \mid \mu \sim \text{Poisson}(\mu)$$

1. Describe the statistical properties of the random variable $\mu$. Specifically, provide its **probability density function**, **expectation**, and **variance**.
2. Derive the **expectation** and **variance** of the times a task has to be performed on a random day, that are $E[X]$ and $Var(X)$ respectively. Justify rigorously all the steps of your derivation. (*Hint: use the law of total expectation and the law of total variance*)

The duration required to execute once the assigned task, denoted as $T$, depends on the server assigned for the day. The system can be allocated one of two servers, each with distinct average processing times ($\tau_1$ and $\tau_2$ respectively).

**Server A**: High-speed processing, average time $\tau_1$, assigned with probability $w$

**Server B**: Standard processing, average time $\tau_2$, assigned with probability $1 - w$

Hence the expectation $\tau$ of the random time $T$ is random, drawn independently each day, and follows a categorical distribution fully defined by:

$$P(\tau = \tau_1) = w$$
$$P(\tau = \tau_2) = 1 - w$$

The random variable $T$ is modelled conditioning on $\tau$ accordingly:

$$T \mid \tau \sim Exp(1/\tau)$$

3. Derive the **expectation** and **variance** of the time needed to process the task once on a random day, that are $E[T]$ and $Var(T)$ respectively. Justify rigorously all the steps of your derivation. (*Hint: use the law of total expectation and the law of total variance*)

**Assume that $X$ and $T$ are independent**.

4. Derive the **expectation** of the time taken by the unit to process all tasks of a single day, i.e. $E[TX]$. Justify rigorously all the steps of your derivation.
5. *Use the law of total variance to show <u>in full generality</u> that, given two independent random variables $T, X$, the following equality holds:

$$Var(TX) = E^2[X]Var(T) + E^2[T]Var(X) + Var(X)Var(T)$$

6. Use the previous result to derive explicitly (in terms of $\alpha, \beta, \tau_1, \tau_2, w$) the **variance** of the time taken by the unit to process all tasks of a single day, i.e. $Var(TX)$. Justify rigorously all the steps of your derivation.

## PART II - Computation

Considering the parameters values $\alpha = 10$, $\beta = 1$, $\tau_1 = 10$ min, $\tau_2 = 25$ min and $w = 0.6$.
You are interested in studying in more depth the distributions of the random variables $X, T$ and $XT$.

1. Calculate expectation and variance of the three random variables.
2. Produce three plots to visualise the (approximated) distribution of the three random variables.
   (*Hint: computing explicitly the distributions in terms of their pdf and cdf might be difficult or not feasible. Use instead stochastic simulation.*
   *Produce a large sample drawn from each random variable. Exploit the hierarchical structure of the model. A histogram of such large sample constitute a reasonable approximation of the respective variable distribution i.e. pdf)*
3. Estimate the probability that the time taken by the unit to process all tasks of a single day is above 5 hours, i.e. $P(TX > 5$ hours).
   (*Hint: approximate this probability using stochastic simulation. First simulate a large sample from $TX$. Then evaluate the proportions of sample above 5 hours*)
4. *Estimate three intervals such that the probability that the time taken by the unit to process all tasks of a single day is in the interval is (approximately) $0.90$, $0.95$ and $0.99$ respectively.
   (*Hint: keep using your simulated sample from $TX$. Find such intervals by looking at the suitable quantiles of the sample*)

Assume the cost per hour to run the unit is $100$€ for the first hour, and increases by $30$€ each hour: the second hour it costs 130€, the third 160€,…
For example, the cost to run the unit for 3 hours is $390$€. Also, the unit cannot be deployed for fraction of an hour, so the cost to run it for e.g. 3 hours is the same as the cost to run it for 2 hours and 30 minutes.

5. *Estimate the expectation and variance of the cost to run the unit on a random day.
6. *Produce a suitable visualisation of the distribution of the cost to run the unit on a random day.

(*) Exercises with grey background and marked with a * are increasingly challenging, contribute very marginally to the score and are considered only if the rest of the assignment is perfect. You can reach 8-9 points out of 10 without them.