Applied Stochastic Processes (30515) - Assignment 1

Andrea Lisci

March 2024

1 Part I

1.1 Description of the random variable μ

The random variable μ represents the expected number of times that the task has to be performed and it is described by a gamma function with parameters α and β ($\Gamma(\alpha, \beta)$). Its PDF is [1]:

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{if } x \geq 0 \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

The expectation of this random variable is:

$$E[\mu] = \frac{\alpha}{\beta} \tag{2}$$

And the variance is:

$$var(\mu) = \frac{\alpha}{\beta^2} \tag{3}$$

1.2 Expectation and variance of the random variable X

In the exercise tho, we are interested in determining how many times a task has to be permormed on a random day, so we are interested in the expectation and variance of the random variable X.

Let's start by deriving the expectation, by using the law of total expectation:

$$E[X] = E[E[X|\mu]]$$

One of our hypoteses says that:

$$X|\mu \sim Poisson(\mu)$$

[1] From: `https://en.wikipedia.org/wiki/Gamma_distribution`

So $E[X|\mu]$ is simply equal to the expectation of a Poisson distribution with parameter μ:

$$E[E[X|\mu]] = E[\mu] = \frac{\alpha}{\beta} \tag{4}$$

We can now calculate the variance, by using the law of total variance we have:

$$var(X) = E[var(X|\mu)] + var(E[X|\mu])$$

Since $X|\mu \sim Poisson(\mu)$, we have that:

$$var(X) = E[\frac{\alpha}{\beta^2}] + var(\frac{\alpha}{\beta}) = \frac{\alpha}{\beta^2} + \frac{\alpha}{\beta} \tag{5}$$

1.3 Expectation and variance of the random variable T

We now define a new random variable called T that represents the random time took to process the task. Each server has a different average processing time (τ_1,τ_2,τ_3 respectively). We also have that:

- $P(\tau = \tau_1) = 0.3$
- $P(\tau = \tau_2) = 0.5$
- $P(\tau = \tau_3) = 0.2$

And:

$$T|\tau \sim Exp(\frac{1}{\tau})$$

It's an exponential, so [2]:

- $E[T|\tau] = 1/1/\tau$
- $var(T|\tau) = 1/1/\tau^2$

So, by using the law of total expectation we have that:

$$E[T] = E[E[T|\tau]] = E[\frac{1}{\frac{1}{\tau}}] = E[\tau]$$

So:

$$E[T] = 0.3\tau_1 + 0.5\tau_2 + 0.2\tau_3 \tag{6}$$

We can now start thinking about the variance. For the law of total variance we have that:

$$var(T) = E[var(T|\tau)] + var(E[T|\tau])$$

Since $T|\tau \sim Exp(\frac{1}{\tau})$, $var(T|\tau) = 1/(1/\tau^2)$, so:

$$var(T) = E[1/(1/\tau^2)] + var(\tau) = E[\tau^2] + var(\tau)$$

By using the definition of variance, we have that:

$$var(\tau) = E[\tau^2] - E^2[\tau]$$

So:

$$var(T) = E[\tau^2] + E[\tau^2] - E^2[\tau] = 2E[\tau^2] - E^2[\tau] \tag{7}$$

[2] From: `https://en.wikipedia.org/wiki/Exponential_distribution`

1.4 Expectation of the random variable TX

We now want to know how the time taken by the unit to process all tasks of a single day behaves, we start from the expectation $E[TX]$. By using the law of total expectation we have that:

$$E[TX] = E[E[TX|X]]$$

Now we can use the definition of measurability ($E[f(x)Y|X] = f(x)E[Y|X]$) to obtain:

$$E[TX] = E[E[TX|X]] = E[XE[T|X]]$$

And now, since T and X are independent we can use the definition of independence (if X and Y are independent $E[Y|X] = E[Y]$):

$$[E[TX] = E[E[TX|X]] = E[XE[T|X]] = E[XE[T]]$$

$E[T]$ is a constant, so $E[XE[T]] = E[X]E[T]$:

$$E[XT] = E[X]E[T] = \frac{\alpha}{\beta}(0.3\tau_1 + 0.5\tau_2 + 0.2\tau_3) \tag{8}$$

1.5 Variance of the random variable TX

At this point we want to analize the variance of the random variable TX, in particular we want to prove, using the law of total variance, that:

$$var(TX) = E^2[X]var(T) + E^2[T]var(X) + var(X)var(T) \tag{9}$$

We start by applying the law of total variance:

$$var(TX) = E[var(TX|X)] + var(E[TX|X])$$

Since $var(TX|X) = X^2 var(T|X) = X^2 var(T)$ and $E[TX|X] = XE[T|X] = XE[T]$for independence and measurability we have this result:

$$var(TX) =$$

$$= E[X^2 var(T)] + var(XE[T])$$

Now, $var(T)$ and $E[T]$ are constants, so:

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

We also know that $var(x) = E[x^2] - E^2[x]$, so:

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

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

1.6 Derivation of $var(TX)$

We can now sobstitute our previous results in the formula for the variance of the random variable TX:

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

By using (4), (5), (6), (7):

$$var(TX) = (\frac{\alpha}{\beta^2} + \frac{\alpha}{\beta})(2E[\tau^2] - E^2[\tau]) + (\frac{\alpha}{\beta})^2(2E[\tau^2] - E^2[\tau]) +$$

$$+E^2[T](\frac{\alpha}{\beta^2} + \frac{\alpha}{\beta})$$

By grouping by $(\frac{\alpha}{\beta^2} + \frac{\alpha}{\beta})$ we have:

$$var(TX) = (\frac{\alpha}{\beta})^2(2E[\tau^2] - E^2[\tau]) + (\frac{\alpha}{\beta^2} + \frac{\alpha}{\beta})(2E[\tau^2]) \tag{10}$$

2 Part II

2.1 Expectation and variance of the three random variables

For this part of the assignment I used Python, by letting $\alpha = 10$, $\beta = 1$, $\tau_1 = 10$, $\tau_2 = 20$, $\tau_3 = 30$ and by applying the equations derived in the previous points, we get:

- $E[X] = 10$
- $E[T] = 19$
- $E[TX] = 190$
- $var(X) = 20$
- $var(T) = 459$
- $var(TX) = 62300$

2.2 Plots for the three random variables

We know want to have some insights on the distributions of the three random variables, the plots of the three distributions are:

- X: Figure 1
- T: Figure 2
- TX: Figure 3

2.3 Probability that the time taken by the unit to process all tasks of a single day is above 5 hours

This probability is equivalent to the problem $P(TX > 5 \text{ hours})$. Since we don't have a clear function that describes the distribution of TX we can count the generated values that are greater than "5 hours" (300 minutes), with this idea we get that:

$$P(TX > 5) = 0.1958$$

approximately (figure 4).

2.4 Confidence intervals

We are now interested in estimating three confidence intervals with 0.9, 0.95, 0.99 probabilities. We can do this by sorting and counting values like we did in the previous point, so (Figure 5):

- $CI_{0.9} = [6.455288079690898, 664.2884257261212]$
- $CI_{0.95} = [3.1089032265444794, 869.1936723469951]$
- $CI_{0.99} = [0.46642474750413904, 1487.0026474631752]$

2.5 Expectation and variance of the cost to run the unit on a random day

Now we also know that the cost to run the unit is 100€ for the first hour, then it increases by 30€ each hour, also the cost to run it for e.g. 3 hours is the same as the cost to run it for 2 hours and 59 minutes. To calculate the expectation and the variance of this cost function we can use a cycle to create a new list where the times under 60 minutes are added as 100, while the other are added as $\sum_{i=0}^{k} 100 + 30i$. After running the code we get that (Figure 6):

- $E[cost] = 786.858$ approximately
- $var(cost) = 3680751.3569551297$ approximately

2.6 Distribution of the cost to run the unit on a random day

Since we already have a list of the costs, we can just plot it. (Figure 7)

In order to show the distribution better, I decided to use a logarithmic scale on the y axis of the plot.

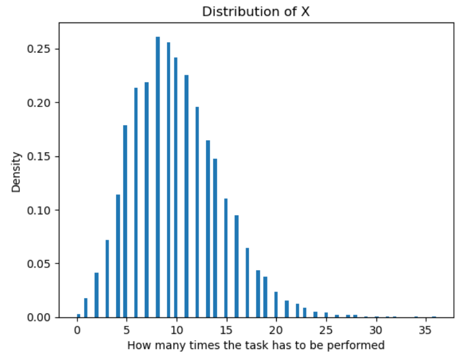

Figure 1: Distribution of X

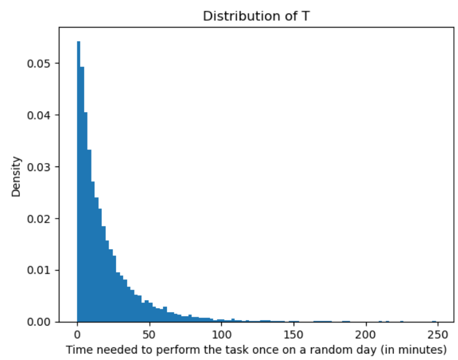

Figure 2: Distribution of T

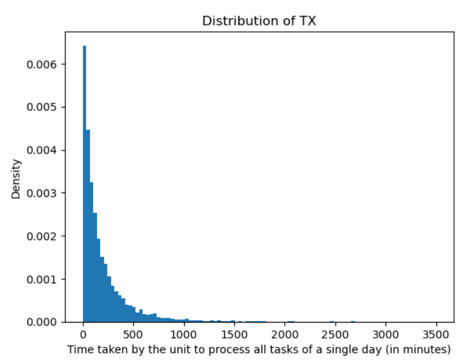

Figure 3: Distribution of TX

```python
count = 0
for s in distr_TX:
    if s > 300:
        count = count + 1
    else:
        count = count
prob = count/len(distr_TX)
prob
```

```
0.1958
```

Figure 4: $P(TX > 5)$

```python
distr_TX.sort(reverse = False)
i = round(len(distr_TX)*0.05)
j = round(len(distr_TX)*0.95)
est1 = distr_TX[i]
est2 = distr_TX[j]
conf_int90=[est1,est2]
conf_int90
```

```
[6.455288079690898, 664.2884257261212]
```

```python
i = round(len(distr_TX)*0.025)
j = round(len(distr_TX)*0.975)
est1 = distr_TX[i]
est2 = distr_TX[j]
conf_int95=[est1,est2]
conf_int95
```

```
[3.1089032265444794, 869.1936723469951]
```

```python
i = round(len(distr_TX)*0.005)
j = round(len(distr_TX)*0.995)
est1 = distr_TX[i]
est2 = distr_TX[j]
conf_int99=[est1,est2]
conf_int99
```

```
[0.46642474750413904, 1487.0026474631752]
```

Figure 5: Confidence intervals

```
cost = []
for w in distr_TX:
    if w <= 60:
        cost.append(100)
    else:
        x = np.ceil((w)/60)
        costo =0
        i = 0
        while i < x:
            costo = costo + 100 + 30*i
            i = i+1
        cost.append(costo)
```

```
expectation_cost = sum(cost)/len(cost)
expectation_cost
```

```
782.3062
```

```
cost_variance_list = []
cost_variance = np.var(cost,ddof = 1)
cost_variance
```

```
3680751.3569551297
```

Figure 6: Expectation and variance of the costs

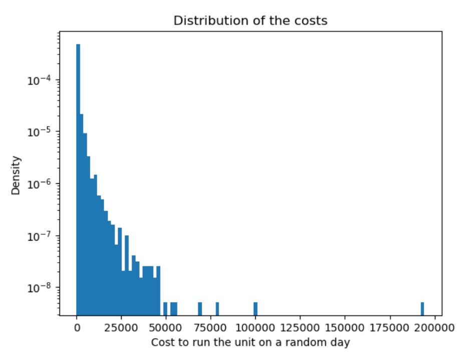

Figure 7: Distribution of the costs (with logarithmic scale on the y axis)