Problem 1

<u>Question a</u>

From the lectures we know that triangle distribution has probability density function.

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & , \quad a \le x \le c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & , \quad c < x \le b \\ 0 & , \quad otherwise \end{cases}$$

From that we can derive cumulative distribution function.

The cdf on $a \le x \le c$

$$F(x) = \int_a^x \frac{2(t-a)}{(b-a)(c-a)}\, dt = \frac{(x-a)^2}{(b-a)(c-a)}$$

The cdf on $c < x \le b$

$$F(x) = \int_c^x \frac{2(b-t)}{(b-a)(b-c)}\, dt = 1 - \frac{(x-b)^2}{(b-a)(b-c)}$$

So, cumulative distribution function will look like:

$$F(x) = \begin{cases} \dfrac{(x-a)^2}{(b-a)(c-a)} & , \quad a \le x \le c \\ 1 - \dfrac{(x-b)^2}{(b-a)(b-c)} & , \quad c < x \le b \\ 0 & , \quad otherwise \end{cases}$$

When equating the cdf to u where $0 \le x \le 1$ yields an inverse cdf.

Case 1 $a \le x \le c$**:**

$$\frac{(x-a)^2}{(b-a)(c-a)} = u$$

$$\frac{x-a}{\sqrt{(b-a)(c-a)}} = \sqrt{u}$$

$$\sqrt{(b-a)(c-a)} \cdot (\sqrt{u}) = x - a$$

$$a + \sqrt{(b-a)(c-a)u} = x$$

Case 2 $c < x \leq b$**:**

$$1 - \frac{(x-b)^2}{(b-a)(b-c)} = u$$

$$-\frac{(x-b)^2}{(b-a)(b-c)} = -1 + u$$

$$\frac{(x-b)^2}{(b-a)(b-c)} = 1 - u$$

$$\frac{x-b}{\sqrt{(b-a)(b-c)}} = \sqrt{1-u}$$

$$\sqrt{(b-a)(b-c)} \cdot \sqrt{1-u} = b - x$$

$$b - \sqrt{(b-a)(b-c)(1-u)} = x$$

$$F^{-1}(u) = \begin{cases} a + \sqrt{(b-a)(c-a)u} & , \; 0 \leq u \leq \dfrac{c-b}{b-a} \\ b - \sqrt{(b-a)(b-c)(1-u)} & , \; \dfrac{c-b}{b-a} < u \leq 1 \end{cases}$$

So, the algorithm for general values a, b and c will be

1) First generate U ~ U (0,1)

2) If $U \leq \frac{c-a}{b-a}$ then $X \leftarrow a + \sqrt{(b-a)(c-a)U}$

3) else $X \leftarrow b - \sqrt{(b-a)(b-c)(1-u)}$

4) return X

Problem 2

<u>Question a</u>

From the information given in the assignment we have $\mu = (75, 46, 18)^T$.

To find the covariance matrix recall from the lectures that

$$p_{12} = Corr(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)Var(X_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$$

$$p_{13} = Corr(X_1, X_3) = \frac{Cov(X_1, X_3)}{\sqrt{Var(X_1)Var(X_3)}} = \frac{\sigma_{13}}{\sigma_1\sigma_3}$$

$$p_{23} = Corr(X_2, X_3) = \frac{Cov(X_2, X_3)}{\sqrt{Var(X_2)Var(X_3)}} = \frac{\sigma_{23}}{\sigma_2\sigma_3}$$

So, we get that

$$\sigma_{12} = p_{12}\sigma_1\sigma_2$$

$$\sigma_{13} = p_{13}\sigma_1\sigma_3$$

$$\sigma_{23} = p_{23}\sigma_2\sigma_3$$

For the first suggested scenario we can calculate the values of the covariance matrix as follows.

$$\sigma_{11} = \sigma_1^2 = 625$$

$$\sigma_{22} = \sigma_2^2 = 100$$

$$\sigma_{33} = \sigma_3^2 = 25$$

$$\sigma_{12} = \sigma_{21} = -0.75 \cdot 25 \cdot 10 = -187.5$$

$$\sigma_{13} = \sigma_{31} = 0 \cdot 25 \cdot 5 = 0$$

$$\sigma_{23} = \sigma_{32} = 0 \cdot 10 \cdot 5 = 0$$

$$\sum = \begin{bmatrix} 625 & -185.5 & 0 \\ -185.5 & 100 & 0 \\ 0 & 0 & 25 \end{bmatrix}$$

For the second suggested scenario we can calculate the values of the covariance matrix as follows.

$$\sigma_{11} = \sigma_1^2 = 625$$

$$\sigma_{22} = \sigma_2^2 = 100$$

$$\sigma_{33} = \sigma_3^2 = 25$$

$$\sigma_{12} = \sigma_{21} = 0.75 \cdot 25 \cdot 10 = 187.5$$

$$\sigma_{13} = \sigma_{31} = 0 \cdot 25 \cdot 5 = 0$$

$$\sigma_{23} = \sigma_{32} = 0 \cdot 10 \cdot 5 = 0$$

$$\sum = \begin{bmatrix} 625 & 185.5 & 0 \\ 185.5 & 100 & 0 \\ 0 & 0 & 25 \end{bmatrix}$$

For the third suggested scenario we can calculate the values for the covariance matrix as follows.

$$\sigma_{11} = \sigma_1^2 = 625$$

$$\sigma_{22} = \sigma_2^2 = 100$$

$$\sigma_{33} = \sigma_3^2 = 25$$

$$\sigma_{12} = \sigma_{21} = -0.75 \cdot 25 \cdot 10 = -187.5$$

$$\sigma_{13} = \sigma_{31} = 0.4 \cdot 25 \cdot 5 = 50$$

$$\sigma_{23} = \sigma_{32} = -0.5 \cdot 10 \cdot 5 = -25$$

$$\sum = \begin{bmatrix} 625 & -185.5 & 50 \\ -185.5 & 100 & -25 \\ 50 & -25 & 25 \end{bmatrix}$$

The scenario with $p_{12} = 0.75$ will most probably generate a top tier character, because with a positive correlation between variables X_1 and X_2 they will be both disposed to have a high value as well as have a low value. That gives a higher chance for a top tier than if we have a negative correlation, where a variable with high value gets together with a low value of the other variable and seldom gives a high value on both variables on the same time.

Problem 3

Question a

For the MC approach we first write the integral as:

$$\theta = 24\int_a^b \lambda(x)\frac{1}{24}dx =$$

$$= 24\int_a^b \left(5 + 50\sin\left(\pi\cdot\frac{x}{24}\right)^2 + 190e^{\frac{-(x-20)^2}{3}}\right)\frac{1}{24}dx =$$

$$= 24\cdot E(5 + 50\sin\left(\pi\cdot\frac{X}{24}\right)^2 + 190e^{\frac{-(X-20)^2}{3}})$$

Which is estimated using:

$$\hat{\theta}_{CMC} = \frac{b-a}{n}\sum_{i=1}^{n} g(X_i)$$

Where $X_1, X_2, \dots, X_n$ $are\ iid\ U[a,b]$.

According to that we will get:

$$\hat{\theta}_{CMC} = \frac{24}{n}\sum_{i=1}^{n}(5 + 50\sin\left(\pi\cdot\frac{X_i}{24}\right)^2 + 190e^{\frac{-(X_i-20)^2}{3}})$$

$\hat{\theta}_{CMC}$ will be normally distributed since it's an average of many observations. So we can state that there is approximately $a(1-\alpha)100\%$ probability that $\hat{\theta}$ falls in :

$$\left[\theta \pm z_{\frac{\alpha}{2}}SD(\hat{\theta})\right]$$

Furthermore, from the lectures example we have that

$$SD\left(\hat{\theta}_{CMC}\right) = (b-a)\sigma_{g(X)}/\sqrt{n} = 24\sigma_{\lambda(X)}/\sqrt{n}$$

I.e. to have $a(1-\alpha)100\%$ probability that $\hat{\theta}$ is at most a margin e from θ we need:

$$z_{\frac{\alpha}{2}} \frac{(b-a)\sigma_{g(X)}}{\sqrt{n}} < e \Rightarrow n > \frac{{z_{\frac{\alpha}{2}}}^2 (b-a)^2 \sigma_{g(X)}^2}{e^2}$$

Where $\sigma_{g(X)}^2$ can be estimated by a first simulation and then plugged into the formula.

So with that we have:

$$z_{\frac{\alpha}{2}} \frac{24\sigma_{\lambda(X)}}{\sqrt{n}} < e \Rightarrow n > \frac{{z_{\frac{\alpha}{2}}}^2 24^2 \sigma_{\lambda(X)}^2}{e^2}$$

To be at least 95% certain that the estimate is no more than 10 from the true answer we can calculate the formula above assuming that $z_{\frac{\alpha}{2}}$ = 1.96

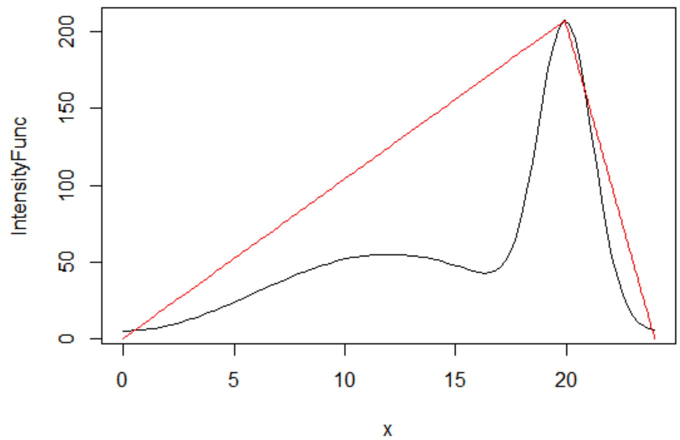
$$n > \frac{1.96^2 24^2 \hat{\sigma}_{\lambda(X)}^2}{10^2}$$

Question c

We cannot use antithetic variables to improve the precision of ordinary Monte Carlo because the function:

$$\int_0^{24} \lambda(t)dt = \int_0^{24} \left(5 + 50\sin\left(\pi \cdot \frac{t}{24}\right)^2 + 190e^{\frac{-(t-20)^2}{3}}\right)dt$$

on the interval [a, b] is not monotonic. This is can be proved by making a plot of the curve in R.



If our function $\lambda(t)$ decomposed into $\lambda(t) = h(t)f(t)$ where $f(t)$ will be the density function (importance function). We can alternatively write $h(t)$ function like $\frac{\lambda(t)}{f(t)}$. In my opinion a good choice of a density function for the importance sampling can be the density function of triangle distribution from a = 0 to b = 24 with c as maximum value of $\lambda(t)$. After calculating the variance from the difference between functions $\lambda(t)$ *and* $h(t)$ we can state that the result is smaller than the MC integration, which means that the precision has improved.

The algorithm for calculation the importance sampling is as follows.

1. Simulate a large number of values from the triangle distribution with given parameters (a,b,c).
2. Calculate the mean of the function $\frac{\lambda(X)}{f(X)}$ and save it to a vector of values Vec. Where $\lambda(x) = \left(5 + 50\sin\left(\pi \cdot \frac{x}{24}\right)^2 + 190e^{\frac{-(x-20)^2}{3}}\right)$, and $f(x)$ is the proposed triangle distribution.
3. Repeat step1 and step2 big number of times.
4. Calculate the standard deviation of importance sampling by calculating the standard deviation of the resulting vector of values Vec.

Problem 4

Question a

To show that the maximum likelihood estimator (MLE) for λ is $\hat{\lambda} = n / \sum_{i=1}^{n} X_i$, where the nucleation time is exponentially distributed with a rate λ called the nucleation rate, we have to follow the next steps.

We have that $X_1, \dots, X_n$ i.i.d. $f(x;\theta)$.

1. Define the likelihood function:

$$L(\lambda) = f(x_1, \dots, x_n; \lambda) =$$

$$= f(x_1;\lambda) \dots f(x_n;\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) =$$

$$= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = (\lambda^n) e^{-\sum_{i=1}^{n} -\lambda x_i}$$

2. Take ln() for simpler optimization of L(θ):

$$l(\lambda) = \ln[L(\lambda)] =$$

$$= \ln\left((\lambda^n) e^{-\sum_{i=1}^{n} -\lambda x_i}\right) =$$

$$= \mathrm{n} \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

3. Take the derivative w.r.t. λ, set equal to 0 and solve:

$$\frac{\partial}{\partial \lambda} l(\lambda) = 0 \Rightarrow \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \lambda = \frac{n}{\sum_{i=1}^{n} x_i} \Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Bootstrap procedure:

1. Repeat step 2-3 for b = $1, 2, \dots, B$.
2. Simulate data $x_1^{*(b)}, \dots, x_n^{*(b)}$ from the estimated distribution $\hat{F} = \hat{\lambda}$ by drawing n observations with replacement from the original data $x_1, \dots, x_n$.

3. Calculate $\hat{\lambda}^{*(b)}$ $from\ x_1^{*(b)}, \ldots, x_n^{*(b)}$.

The ecdf of $\hat{\lambda}^{*(1)}, \ldots, \hat{\lambda}^{*(B)}$ will be the estimate of the distribution of $\hat{\lambda}$.

The ***standard deviation*** of the estimator $\hat{\lambda}$ is estimated by the SD of $\hat{\lambda}^{*(1)}, \ldots, \hat{\lambda}^{*(B)}$:

$$\widehat{SD}(\hat{\lambda}) = \widehat{SD}(\hat{\lambda}^{*}) = \sqrt{\frac{1}{n-1}\sum_{b=1}^{B}(\hat{\lambda}^{*(b)} - \overline{\hat{\lambda}^{*}})^2}$$

Where $\overline{\hat{\lambda}^{*}} = \sum_{b=1}^{B}\frac{\hat{\lambda}^{*(b)}}{B}$.

Bootstrap estimate for the ***bias*** the estimator:

$$\widehat{bias}(\hat{\lambda}) = \overline{\hat{\lambda}^{*}} - \hat{\lambda}$$

Where $\overline{\hat{\lambda}^{*}} = \sum_{b=1}^{B}\frac{\hat{\lambda}^{*(b)}}{B}$.

There are several ways the bootstrap approach can be used to obtain approximate confidence intervals. One of the possible ways to calculate approximate confidence intervals for λ is called "Standard normal bootstrap interval". Which is a simple interval but often not the best. Generally SD($\hat{\lambda}$) is unknown, but can be estimated by the bootstrap standard deviation estimate to obtain the interval:

$$[\hat{\lambda} \pm z_{a/2}\widehat{SD}(\hat{\lambda}^{*})]$$

<u>Question c</u>

Permutation hypothesis test can be constructed as follows:

1. Compute the test statistics $T = T(X,Y)$ for original data
2. Repeat steps 3-4 for a number $p = 1, 2, \ldots, P$ of time.

3. Simulate data $X^{*(p)} = \left(X_1^{*(p)}, \dots, X_{nX}^{*(p)}\right) = \ Y^{*(p)} = \left(Y_1^{*(p)}, \dots, Y_{nY}^{*(p)}\right)$ by drawing $n_X + \ n_Y$ observations without replacement from the original data $X_1, \dots, X_nX, \ Y_1, \dots, Y_nY.$ I.e. permuting the order of original data.
4. Calculate $T^{*(p)} = T(X^{*(p)}, Y^{*(p)})$
5. Calculate $p - value = \ \sum_{i=1}^{P} I(|X^{*(p)}| > |T|)/P$. If large absolute values of T support the alternative hypothesis. Otherwise adjust the $p - vlaue$ calculation accordingly.
6. Reject the null hypothesis if p-value < α, where α = 0.05.

Since our data is not normally distributed, we have to construct a permutation test based on the test statistic, which will be the difference of the means.

$$H_0 : \ \mu_X = \mu_Y \quad \text{vs} \quad H_1 : \ \mu_X \neq \mu_Y$$

The permutation test will be based on:

$$\hat{\mu}_{diff} = \ \hat{\mu}_X - \hat{\mu}_Y = \overline{X} - \overline{Y}$$

Where $\overline{X}\ and\ \overline{Y}$ are the means of measurements in X and Y respectively.

We also can perform another test statistic, which will be based on the difference of the estimators.

$$H_0 : \ \hat{\lambda}_X = \hat{\lambda}_Y \quad \text{vs} \quad H_1 : \ \hat{\lambda}_X \neq \hat{\lambda}_Y$$

The permutation test will be based on:

$$\hat{\lambda}_{diff} = \ \hat{\lambda}_X - \hat{\lambda}_Y$$

<u>Question d</u>

So, having that p-value is smaller in each test than $\alpha = 0.05$ we can reject H_0 and state that $\mu_X \neq \mu_Y$ and $\hat{\lambda}_X \neq \hat{\lambda}_Y$.

Question e

For the hypothesis test to test that $\lambda > 0.0003$ we specify two different hypothesis:

1) $H_0: \lambda = 0.0003$
2) $H_1: \lambda > 0.0003$

An algorithm to specify the test statistic:

1) Generate a large number (N) of vectors with random exponential distribution, with the assumption that H_0 is correct.
2) For each vector solve $\hat{\lambda}^{*(p)} = \frac{n}{\sum_{i=1}^{n} X_i}$, where n is the length of the vector, and X_i an element inside the vector and $p = 1,2,3, \ldots, N$.
3) Use the obtained set $T^{*(p)}$ with a new lambda that is equal to the number of simulations to find $p - value$ using this formula.
   $p - value = \sum_{p=1}^{P} I\left(T^{*(p)} > \hat{\lambda}\right)/P$

No, we could not alternatively construct a permutation test because the permutation test requires two different datasets obtained under similar conditions to make use of the algorithm, but we only have one.

Problem 5

Question a

Based on the function given in the assignment:

$$g(s) = \begin{cases} 0 & ,when\ \ 0 \le s \le t_0 \\ \frac{\beta}{t_s}(s - t_0) & ,when\ \ t_0 \le s \le t_0 + t_s \\ \beta & ,when\ \ t_0 + t_s \le s \le t_0 + t_s + t_p \\ \beta e^{-\gamma\left(s-\left(t_0+t_s+t_p\right)\right)} & ,when\ \ s > t_0 + t_s + t_p \end{cases}$$

To show that the total production up to time s will be:

$$G(s) = \int_0^s g(u)du =$$

$$= \begin{cases} 0 & ,when\ \ 0 \leq s \leq t_0 \\ \dfrac{\beta}{2t_s}(s - t_0)^2 & ,when\ \ t_0 \leq s \leq t_0 + t_s \\ \dfrac{\beta}{2}t_s +\ \beta(s - (t_s + t_0)) & ,when\ \ t_0 + t_s \leq s \leq t_0 + t_s + t_p \\ \dfrac{\beta}{2}t_s +\ \beta t_p + \dfrac{\beta}{\gamma}[1 - e^{-\gamma\left(s-\left(t_0+t_s+t_p\right)\right)}] & ,when\ \ s > t_0 + t_s + t_p \end{cases}$$

we must solve the following integrals.

<u>Case 1:</u>

$$G(s) = \ \int_0^s 0du = 0$$

<u>Case 2:</u>

$$G(s) = \int_0^{t_0} 0du + \ \int_{t_0}^s \frac{\beta}{t_s}(u - t_0)du =$$

$$= 0 + \frac{\beta}{t_s}\int_{t_0}^s (u - t_0)du =$$

$$= \frac{\beta}{t_s}\left[\frac{u^2}{2} - t_0 u\right]_{t_0}^s \ = \ \frac{\beta}{t_s}\left[\left(\frac{s^2}{2} - t_0 s\right) - \left(\frac{t_0^2}{2} - t_0^2\right)\right] =$$

$$= \frac{\beta}{2t_s}(s^2 - 2t_0 s - t_0^2) =$$

$$= \frac{\beta}{2t_s}(s - t_0)^2$$

Case 3:

$$G(s) = \int_0^{t_0} 0du + \int_{t_0}^{t_0+t_s} \frac{\beta}{t_s}(u - t_0)du + \int_{t_0+t_s}^{s} \beta du =$$

$$= 0 + \frac{\beta}{t_s}\int_{t_0}^{t_0+t_s}(u - t_0)du + \beta\int_{t_0+t_s}^{s} du =$$

$$= \frac{\beta}{t_s}\left[\frac{u^2}{2} - t_0 u\right]_{t_0}^{t_0+t_s} + \beta[u]_{t_0+t_s}^{s} =$$

$$= \frac{\beta}{t_s}\left[\left(\frac{(t_0+t_s)^2}{2} - t_0(t_0+t_s)\right) - \left(\frac{t_0^2}{2} - t_0^2\right)\right] + \beta[(s) - (t_0+t_s)] =$$

$$= \frac{\beta}{t_s}\left[\left(\left(\frac{t_0^2 + t_s^2 + 2t_0t_s}{2}\right) - (t_0^2 + t_0t_s)\right) - \left(-\frac{t_0^2}{2}\right)\right] + \beta(s - (t_0+t_s)) =$$

$$= \frac{\beta}{2t_s}[(t_0^2 + t_s^2 + 2t_0t_s) - 2(t_0^2 + t_0t_s) + t_0^2] + \beta\big(s - (t_0+t_s)\big) =$$

$$= \frac{\beta}{2}t_s - \beta\big(s - (t_0+t_s)\big)$$

Case 4:

$$G(s) = \int_0^{t_0} 0du + \int_{t_0}^{t_0+t_s} \frac{\beta}{t_s}(s - t_0)du + \int_{t_0+t_s}^{t_0+t_s+t_p} \beta du + \int_{t_0+t_s+t_p}^{s} \beta e^{-\gamma\left(s-(t_0+t_s+t_p)\right)}du$$

$$= 0 + \frac{\beta}{t_s}\left[\frac{u^2}{2} - t_0u\right]_{t_0}^{t_0+t_s} + [\beta u]_{t_0+t_s}^{t_0+t_s+t_p} + \frac{\beta}{\gamma}\left[e^{-\gamma\left(u-(t_0+t_s+t_p)\right)}\right]_{t_0+t_s+t_p}^{s} =$$

$$= \frac{\beta}{t_s}\left[\frac{(t_0+t_s)^2}{2} - t_0(t_0+t_s) - \frac{t_0^2}{2} - t_0\right] + \beta\left[t_0 + t - s + t_p - t_0 - t_s\right] -$$

$$-\frac{\beta}{\gamma}\left[e^{-\gamma\left(s-(t_0+t_s+t_p)\right)} - e^{-\gamma\left(t_0+t_s+t_p-(t_0+t_s+t_p)\right)}\right] =$$

$$= \frac{\beta}{t_s}\left[\frac{t_0^2 + 2t_0t_s + t_s^2}{2} - t_0^2 - t_0t_s - \frac{t_0}{2} - t_0^2\right] + \beta t_p - \frac{\beta}{\gamma}\left[e^{-\gamma\left(s-(t_0+t_s+t_p)\right)} - 1\right] =$$

$$\frac{\beta}{2t_s}[t_0^2 + 2t_0t_s + t_s^2 - 2t_0^2 - 2t_0t_s - t_0^2 + 2t_0^2] + \beta t_p - \frac{\beta}{\gamma}\left[e^{-\gamma\left(s-(t_0+t_s+t_p)\right)} - 1\right] =$$

$$= \frac{\beta t_s^2}{2t_s} + \beta t_p + \frac{\beta}{\gamma}\left[1 - e^{-\gamma\left(s-(t_0+t_s+t_p)\right)}\right] =$$

$$= \frac{\beta}{2}t_s + \beta t_p + \frac{\beta}{\gamma}\left[1 - e^{-\gamma\left(s-(t_0+t_s+t_p)\right)}\right]$$

Question c

The algorithm to simulate the uncertainty in the total production up to time s can be described as follows:

For any of time s from our available times set (5,10,15).

1. Simulate value t_0 with triangle distribution using the following parameters: a = 0.85, b = 1.5, c = 1.1.
2. Simulate value t_s with triangle distribution using the following parameters: a = 0.7, b = 1.7, c = 1.
3. Simulate value t_p with triangle distribution using the following parameters: a = 4, b = 7, c = 5.
4. Simulate value β with triangle distribution using the following parameters: a = 7.5, b = 8.5, c = 8.
5. Simulate value γ with triangle distribution using the following parameters: a = 0.15, b = 0.3, c = 0.25.
6. Calculate the total production function using the parameters specified above: $(s, t_0, t_s, t_p, \beta, \gamma)$ and save the value in a vector.
7. Repeat steps 1-6 a large (N) amount of times.
8. Calculate the summary of the values in the vector.

The minimum number of simulations required to be able to estimate the probability of the 15 years total production to exceed 80 with an error of at most 0.02 with 95% certainty can be calculated with the formula taken from the lectures.

$$n > \frac{1}{e^2}$$

where e is the error of (0.02).

$$\frac{1}{0.02^2} = 2500$$

the minimum amount of required simulations is about 2500.

To calculate the probability of 15 years total production to exceed 80 we can use the formula:

$$p = \frac{Values > 80}{Nsim}$$

Question e

To calculate the threshold production level for given values of $t_0, t_s, t_p, \beta, \gamma$ when the production rate drops below 1, we must solve the following equations:

$$\beta e^{-\gamma\left(s-(t_0+t_s+t_p)\right)} = 1$$

$$e^{-\gamma\left(s-(t_0+t_s+t_p)\right)} = \frac{1}{\beta}$$

$$-\gamma\left(s-(t_0+t_s+t_p)\right) = \ln\left(\frac{1}{\beta}\right)$$

$$s-(t_0+t_s+t_p) = -\frac{1}{\gamma}\ln\left(\frac{1}{\beta}\right)$$

$$s = t_0+t_s+t_p-\frac{1}{\gamma}\ln\left(\frac{1}{\beta}\right)$$

With the equation solved above it is possible to set up a simulation algorithm to simulate the uncertainty in the time to threshold production level and the uncertainty in the total volume production until threshold level reached.

1. Generate the values of $t_0, t_s, t_p, \beta, \gamma$ with triangular distribution.
2. Calculate the threshold using the equation above to find s and save it in a vector.
3. Calculate the total volume production using the value s from step 3 and save it in a vector.
4. Repeat steps 1-4 a large number of times.

To calculate the required number of simulations to be able to estimate the expected time to threshold production level with an error of at most 1 month with 95% certainty, and the required numbers of simulations to be able to estimate the expected total volume produced until the threshold level with an error of at most 0.25 with 95% certainty, we have to use the following equation:

$$\frac{4\sigma^2}{e^2} < n$$

Where e is the error and the σ^2 is the variance of the vector calculated for the thresholds and the total volume produced.

The error in the first case will be $\frac{1}{12} = 0.083$