Mandatory Assignment 2

STK3405 – HØST 19

Fábio Rodrigues Pereira

fabior @ math. uio. no

Problem 1a: why $Y = e^{\sigma U + \mu}$ implies $Y \sim \text{lognorm}(\mu, \sigma)$

- suppose that X is a normal stochastic variable with μ_X and σ_X. Then, it can $\boxed{\text{standardized}}$ by:

$\hookrightarrow U \sim N(0,1)$

$$U = \frac{X - \mu_X}{\sigma_X} \longrightarrow \boxed{X = U \cdot \sigma_X + \mu_X}$$

- we know that $Y = e^{\sigma U + \mu}$, Then:

$$\ln(Y) = \sigma_X U + \mu_X = X$$

- where:
  - $\mu_X :=$ The mean of X;
  - $\sigma_X :=$ the standard deviation of X;
  - $U :=$ The standardized form of X.

$\hookrightarrow$ Then $\ln(Y) = X$ which implies by lognormal definition that a stochastic variable Y is lognormal distributed with parameters μ_X and σ_X if $X = \ln(Y)$ is normally distributed with μ_X and σ_X. ∎

Problem 1b: assume $Y \sim \text{lognormal}(\mu, \sigma)$

let y_α: $P(Y \leq y_\alpha) = \alpha\%$ where $0 \leq \alpha \leq 100$

show $y_\alpha = e^{\sigma u_\alpha + \mu}$ where u_α is the $\alpha\%$ in the standard normal distribution

- assume $Y \sim \text{lognormal}(\mu, \sigma)$, $Y = e^{\sigma U + \mu}$ or $\ln(Y) = \sigma U + \mu$,

hence:

$$\alpha\% = P(Y \leq y_\alpha)$$

$$\alpha\% = P(\ln(Y) \leq \boxed{\ln(y_\alpha)})$$

$$\alpha\% = P(\sigma U + \mu \leq \boxed{\sigma u_\alpha + \mu})$$

then: $\ln(y_\alpha) = \sigma u_\alpha + \mu$

$$y_\alpha = e^{\sigma u_\alpha + \mu}$$

■

Problem 1c: Show $\frac{y_{90}}{y_{50}} = \frac{y_{50}}{y_{10}}$

From 1b, we know that $y_\alpha = e^{\sigma \mu_\alpha + \mu}$, then:

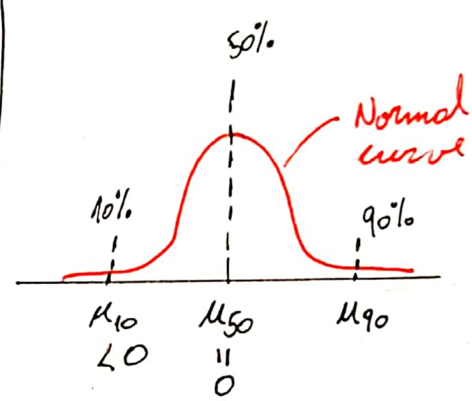
$$\frac{e^{\sigma\cdot\mu_{90}+\mu}}{e^{\sigma\cdot\mu_{50}+\mu}} = \frac{e^{\sigma\cdot\mu_{50}+\mu}}{e^{\sigma\cdot\mu_{10}+\mu}}$$

$$\frac{e^{\sigma\cdot\mu_{90}}\cdot\cancel{e^{\mu}}}{e^{\sigma\cdot\mu_{50}}\cdot\cancel{e^{\mu}}} = \frac{e^{\sigma\cdot\mu_{50}}\cdot\cancel{e^{\mu}}}{e^{\sigma\cdot\mu_{10}}\cdot\cancel{e^{\mu}}}$$

$$e^{\sigma\cdot\cancelto{-\mu_{10}}{\mu_{90}}+\sigma\mu_{10}} = e^{2\sigma\cdot\cancelto{0}{\mu_{50}}}$$

$$e^{0} = e^{0} = 1$$

∎



↳ by symmetry, $\mu_{90} = -\mu_{10}$

Problem 1d: assume $y_{10} = 5$ and $y_{90} = 20$.

Find $y_{50} = ?$

From 1c, we have $\dfrac{y_{90}}{y_{50}} = \dfrac{y_{50}}{y_{10}}$, then $\dfrac{20}{y_{50}} = \dfrac{y_{50}}{5}$

$$\therefore y_{50} = 10$$

Determine μ and σ for this distribution. Assume $\mu_{90} = 1{,}28155$

$$y_{90} = e^{\sigma \cdot \mu_{90} + \mu} \longrightarrow \boxed{20 = e^{\sigma \cdot 1{,}28155 + \mu}}$$

$$y_{50} = e^{\sigma \cdot \mu_{50} + \mu} \longrightarrow 10 = e^{\sigma \cdot 0 + \mu} \longrightarrow 10 = e^{\mu} \rightarrow \boxed{\mu = \ln(10)}$$

$$\therefore \mu = 2.30259$$

Then $20 = e^{\sigma \cdot 1{,}28155} \cdot e^{\ln(10)}$

$$20 = e^{\sigma \cdot 1{,}28155} \cdot 10$$

$$e^{\sigma \cdot 1{,}28155} = 2$$

$$\sigma = \frac{\ln(2)}{1{,}28155} = 0{,}54087$$

Problem – 1.e)

It was created a node called Y with the expression $e^{U*\sigma+\mu}$, where the Y is the lognormal stochastic variable set, U is the standardized normal stochastic variable set, $\sigma = 0{,}54087$ is the standard deviation parameter and $\mu = 2{,}30259$ is the mean/expectation parameter, both parameters from the $Y \sim lognormal(\mu, \sigma)$ distribution, which gave us 10000 different $y_{\alpha\%}$ after simulation and its results are plotted in the S-Curve below:

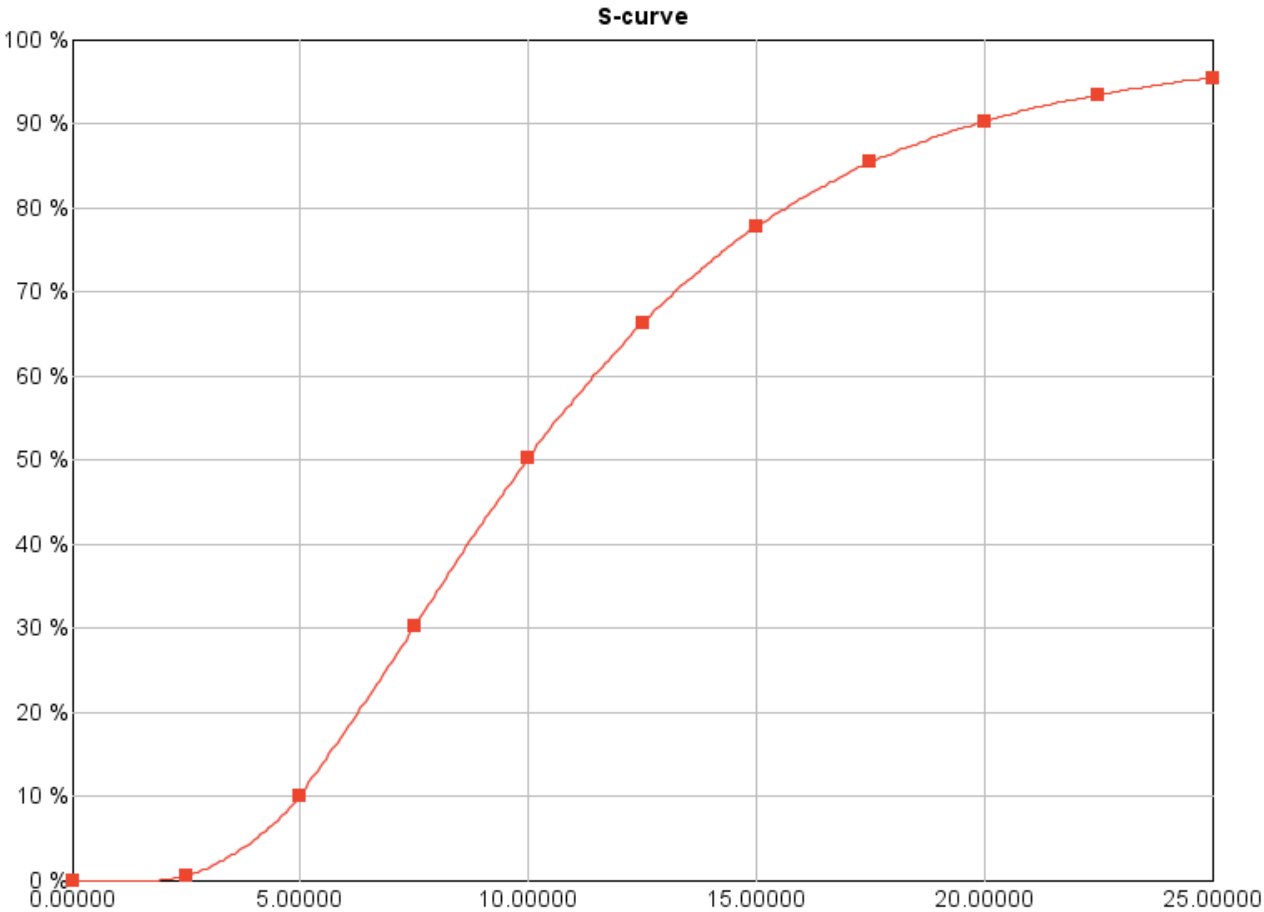


Figure 1. The S-Curve from Riscue Software.

The S-curve shows the cumulative distribution function of the lognormal stochastic variable Y evaluated by $y_{\alpha\%}$ or $P(Y \leq y) = \alpha\%$, which is the probability $\alpha\%$ (in the y-axis) that Y will take a value less than or equal to y (in the x-axis). As expected, the software gave the following results: $y_{10\%} = 5$, $y_{50\%} = 10$ and $y_{90\%} = 20$.

On the other point of view, by the density plot showed below, it is possible to visualize the lognormal equation curve which is positive skewed and with the probabilities of occurrence of each y value (in the x-axis) determined by the area under this curve from minus infinity to y value.

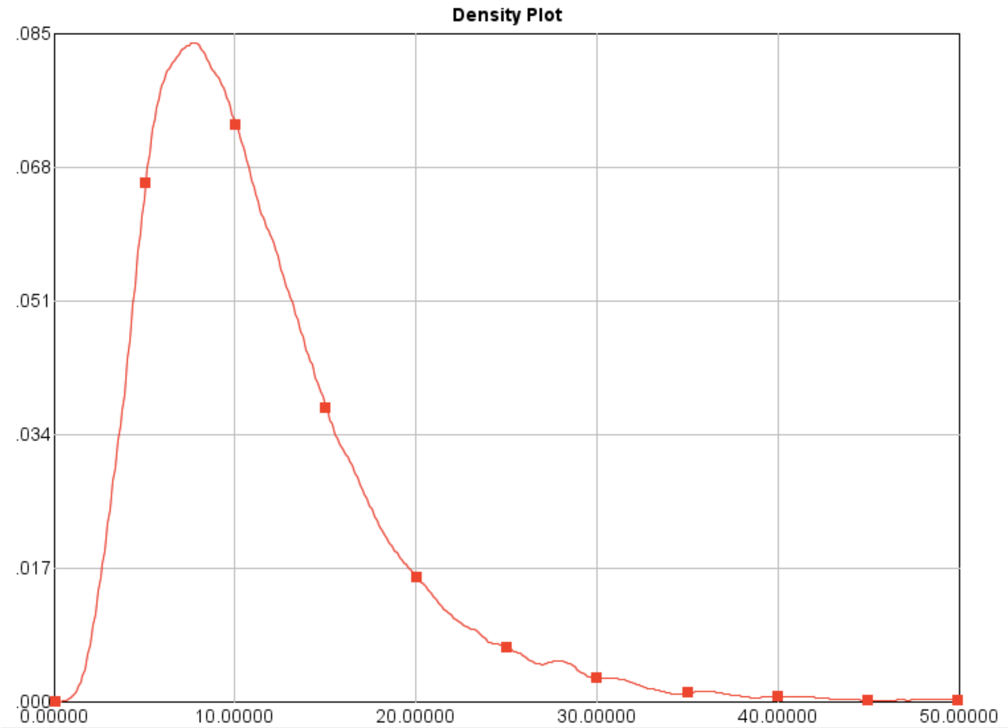


Figure 2. The lognormal density plot from Riscue Software.

Finally, the convergence of the percentiles $y_{90\%}$ *in red*, $y_{50\%}$ *in green and* $y_{10\%}$ *in blue* (in the y-axis) during the 10000 simulations (in % in the x-axis), as showed below:

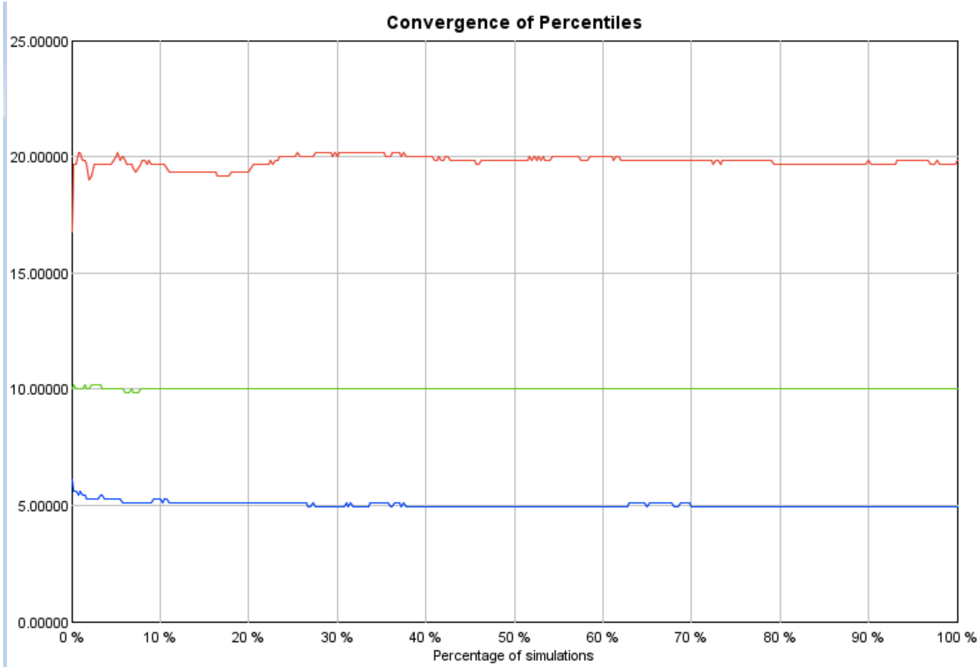


Figure 3. The Convergence of percentiles from Riscue Software.

Here it is showed the speed at which a convergent sequence, in this case the percentiles $y_{\alpha\%}$, approaches its limit. It can be observed that the $y_{90\%}$ approaches its limit at the value 20, the $y_{50\%}$ at 10 and the y_{10} at 5.

Exercise 1f:

$c = cov(U_n, U_{n-1}) = \frac{E\left[(U_n - \mathring{\hat{\mu}})(U_{n-1} - \mathring{\hat{\mu}})\right]}{\cancel{\sigma \cdot \sigma} \to 1} = E\left[U_n . U_{n-1}\right]$

$\boxed{U \sim N(0,1)}$

$E\left[a . U_{n-1}^2 + b . W_n . U_{n-1}\right] = c$ (**)

or

$E\left[\frac{U_n^2 - b . W_n . U_n}{a}\right] = c$ (*)

or

$E\left[U_n \cdot U_{n-1}\right] = c$

or

$E\left[(a . U_{n-1} + b . W_n) \cdot \left(\frac{U_n - b W_n}{a}\right)\right] = c$

$E[X] = \sum_{i=1}^{n} x_i p_i$

or

$E[X] = \int x . f(x)\, dx$

$Var[X] = E[X^2] - E[X]^2$

if $X = c$, then for some $c \in (-\infty, \infty)$ $E[X] = c$

and

$E\left[E[X]\right] = E[X]$

(*) $U_n^2 - b W_n U_n = a . c$ for $|c| \leq 1$

$1 - b W_n U_n = a . c \longrightarrow \frac{1 - b}{a} \leq |c|$

(**) $a . 1 + b E\left[U_{n-1} W_n\right] = c$

$\frac{-1+b}{a} < c < \frac{1-b}{a}$

$b - 1 < c . a < 1 - b$

I Don't know how to solve this. Please, could I get the resolution?

Tusen Takk!

Problem - 1.g)

First, a parameter $c = 0{,}5$ was created to express the correlation between successive terms. Additionally, 2 functions for the constants a and b was created where $a = c$ and $b = \sqrt{1 - a^2}$. Moreover, a chain of $n = 12$ nodes called U_n was created, where $U_1$is obtained by the standard normal distribution with $\mu = 0\ and\ \sigma = 1$ and the others $U_n\ (for\ n = 2, \dots, 12)$, are obtained by $aU_{n-1} + bW_n$ where $W_n \sim N(0, 1)$. Finally, 10000 simulations for each $U_n\ (for\ n = 1, \dots, 12)$ were performed and with the results a correlation table was generated:

Name	U01	U02	U03	U04	U05	U06	U07	U08	U09	U10	U11	U12
U01	1.000	0.494	0.229	0.104	0.053	0.030	0.002	0.010	0.014	0.009	-3.263E-4	0.009
U02	0.494	1.000	0.496	0.249	0.117	0.062	0.034	0.023	0.015	0.020	0.014	0.026
U03	0.229	0.496	1.000	0.508	0.244	0.114	0.069	0.037	0.015	0.005	0.013	0.016
U04	0.104	0.249	0.508	1.000	0.491	0.231	0.118	0.055	0.011	0.016	0.001	0.013
U05	0.053	0.117	0.244	0.491	1.000	0.496	0.247	0.118	0.054	0.039	0.011	0.003
U06	0.030	0.062	0.114	0.231	0.496	1.000	0.501	0.257	0.142	0.077	0.033	0.013
U07	0.002	0.034	0.069	0.118	0.247	0.501	1.000	0.509	0.262	0.132	0.073	0.040
U08	0.010	0.023	0.037	0.055	0.118	0.257	0.509	1.000	0.506	0.263	0.123	0.069
U09	0.014	0.015	0.015	0.011	0.054	0.142	0.262	0.506	1.000	0.511	0.242	0.134
U10	0.009	0.020	0.005	0.016	0.039	0.077	0.132	0.263	0.511	1.000	0.503	0.258
U11	-3.263E-4	0.014	0.013	0.001	0.011	0.033	0.073	0.123	0.242	0.503	1.000	0.499
U12	0.009	0.026	0.016	0.013	0.003	0.013	0.040	0.069	0.134	0.258	0.499	1.000

Figure 4. Correlation table from Riscue software.

The scope of the table above is to verify the correlation between any pair of nodes, but here it is focused only on the pair of consecutives nodes which should have a correlation result of 0,5. As contained in the table, any pair of consecutive node got as result $Corr(U_1, U_2) = 0{,}494$; $Corr(U_2, U_3) = 0{,}496$; $Corr(U_3, U_4) = 0{,}508$; $Corr(U_4, U_5) = 0{,}491$; $Corr(U_5, U_6) = 0{,}496$; $Corr(U_6, U_7) = 0{,}501$; $Corr(U_7, U_8) = 0{,}509$; $Corr(U_8, U_9) = 0{,}506$; $Corr(U_9, U_{10}) = 0{,}511$; $Corr(U_{10}, U_{11}) = 0{,}503\ and\ Corr(U_{11}, U_{12}) = 0{,}499$ which are very approximate to the 0,5 correlation expected. This means that the consecutive nodes are associated with each other and have a measured positive dependence relationship. This is not a surprise, because the expressions of the nodes above represent a time series in which the later node is resulted by a linear combination of the previous node. Furthermore, it is important to notice that the correlation between pairs of consecutive variables are stronger than any other pairs of variables and as far one variable gets from the other as weaker is their correlation, which converge to 0, because $Corr\ =\ 0$ means that the variables have not, or very little, dependence relationship.

Problem - 1.h)

The Riscue software is used this time to transform the time series $\{U_n\}$ above in a new time series $\{Y_n\}$ where the variables are lognormal distributed but not identically. A new row of 12 nodes called Y_n was created below the first 12 nodes called U_n. All the new nodes use the formula $e^{\sigma_n u_n + \mu_n}$, where their $\mu_n and\ \sigma_n$ were gotten from the figures 5 and 6 below:

$$\left.\begin{aligned} \ln(y_{90}) &= \sigma_m \cdot 1{,}28155 + \mu_m \\ \ln(y_{10}) &= \sigma_m \cdot (-1{,}28155) + \mu_m \end{aligned}\right\} \rightarrow \boxed{\mu_m = \ln(y_{90}) - \sigma_m \cdot 1{,}28155}$$

$$\ln(y_{10}) = -1{,}28155\sigma_m + \boxed{\ln(y_{90}) - 1{,}28155\,\sigma_m}$$

$$\ln(y_{10}) - \ln(y_{90}) = -2{,}571\,\sigma_m$$

$$\boxed{\sigma_m = \frac{\ln(y_{10}) - \ln(y_{90})}{-2{,}571}}$$

n	10%	90%
1	5.0	8.2
2	8.0	11.0
3	10.0	12.5
4	10.0	12.5
5	10.0	12.5
6	8.0	11.0
7	6.5	10.0
8	5.0	8.2
9	3.3	7.1
10	2.5	5.5
11	2.1	4.6
12	1.7	3.1

Table 1: Percentiles of $Y_1, \ldots, Y_{12}$.

Figure 5. It shows how the μ_n and σ_n were calculated resulting the next figure.

n	$y_{10\%}$	$y_{90\%}$	μ_n	σ_n
1	5,0	8,2	1,8575	0,1924
2	8,0	11,0	2,2392	0,1239
3	10,0	12,5	2,4145	0,0868
4	10,0	12,5	2,4145	0,0868
5	10,0	12,5	2,4145	0,0868
6	8,0	11,0	2,2392	0,1239
7	6,5	10,0	2,0879	0,1676
8	5,0	8,2	1,8575	0,1924
9	3,3	7,1	1,5782	0,2980
10	2,5	5,5	1,3117	0,3067
11	2,1	4,6	1,1352	0,3050
12	1,7	3,1	0,8319	0,2337

Figure 6 . Percentiles of $Y_1, \ldots, Y_{12}, \mu_n$ and σ_n.

Then, 10000 simulations for the new time series $\{Y_n\}$ were run and the following correlation table generated:

Name	Y01	Y02	Y03	Y04	Y05	Y06	Y07	Y08	Y09	Y10	Y11	Y12
Y01	1.000	0.500	0.244	0.115	0.052	0.029	0.013	0.006	0.001	-0.003	0.001	0.002
Y02	0.500	1.000	0.487	0.246	0.127	0.063	0.035	0.015	0.002	-0.003	-0.003	0.002
Y03	0.244	0.487	1.000	0.495	0.246	0.124	0.075	0.029	0.002	0.009	0.003	0.006
Y04	0.115	0.246	0.495	1.000	0.498	0.244	0.122	0.053	0.022	6.723E-4	0.008	0.014
Y05	0.052	0.127	0.246	0.498	1.000	0.495	0.248	0.125	0.061	0.027	0.022	0.020
Y06	0.029	0.063	0.124	0.244	0.495	1.000	0.503	0.253	0.119	0.059	0.020	0.025
Y07	0.013	0.035	0.075	0.122	0.248	0.503	1.000	0.497	0.245	0.128	0.052	0.030
Y08	0.006	0.015	0.029	0.053	0.125	0.253	0.497	1.000	0.502	0.252	0.123	0.071
Y09	0.001	0.002	0.002	0.022	0.061	0.119	0.245	0.502	1.000	0.486	0.240	0.133
Y10	-0.003	-0.003	0.009	6.723E-4	0.027	0.059	0.128	0.252	0.486	1.000	0.494	0.250
Y11	0.001	-0.003	0.003	0.008	0.022	0.020	0.052	0.123	0.240	0.494	1.000	0.488
Y12	0.002	0.002	0.006	0.014	0.020	0.025	0.030	0.071	0.133	0.250	0.488	1.000

Figure 7. Correlation table from Riscue software.

Once again, the target of this table is to identify measures of dependence relations, positive or negative correlations, between the variables. In this experiment, the correlation between pairs of consecutive variables $Y_n \sim lognormal(\mu, \sigma)$ remains stable on approximately 0,5, which means that they are associated and have a positive dependence correlation. As observed in the exercise 1g, this simulation here has also the same comments regarding the correlation strength.

Problem - 1.i)

In this exercise, the time series $\{Y_n\}$ will be considered as the yearly production of oil (in millions of barrel) from an oilfield with "economic life" of 12 years. For that, it was created 3 new parameters for this statistic model in Riscue, the oil price, exchange rate and discount rate, all with the following respectively default values, 18 dollars per barrel, 7.7 NOK per USD and 7%.

Moreover, 3 new nodes were created, "disc. income", "disc. cost" and "net result" where:

- "Disc. income" is expressed as the product of tree quantities:
    1. First, the present value of the yearly production $\{Y_n\}$, in dollars, which is applied the parameter of the discount rate of 7%;
    2. Next, the parameter oil price, 18usd, which results the gross profit in dollars;
    3. Then, the parameter exchange rate of 7.7, which converts the gross profit from USD to NOK currency.

- "Disc. cost" is acquired from a normal distribution of $\mu = 6000 \; and \; \sigma = 1000$;

- "Net result" is the difference between "Disc. income" and "Disc. cost".

Then, a sensitive test is run by 10000 simulations for the "Net result", and the s-curve and density plots are visualized, as following:

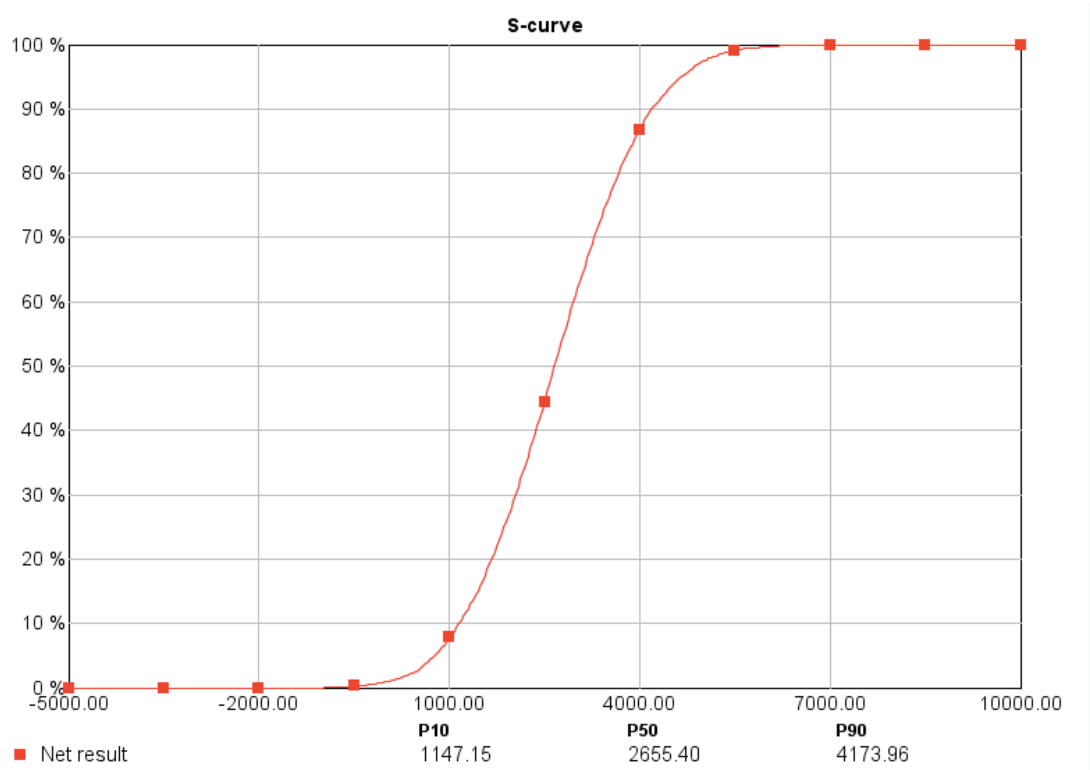


Figure 8. S-curve plot from Riscure software

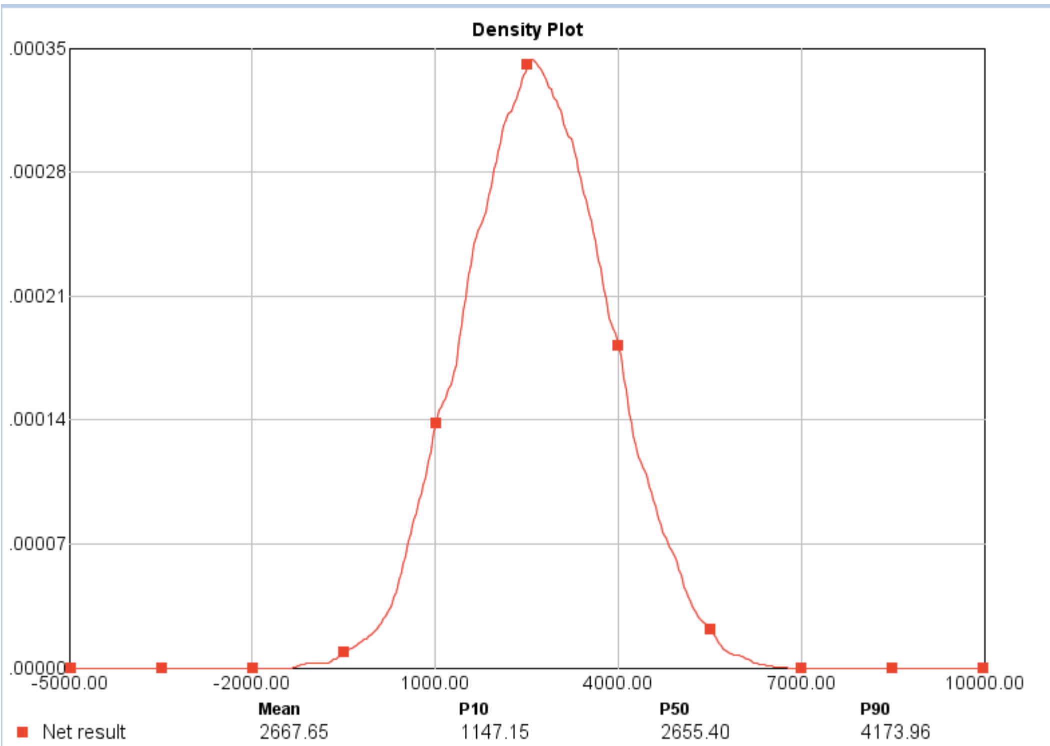
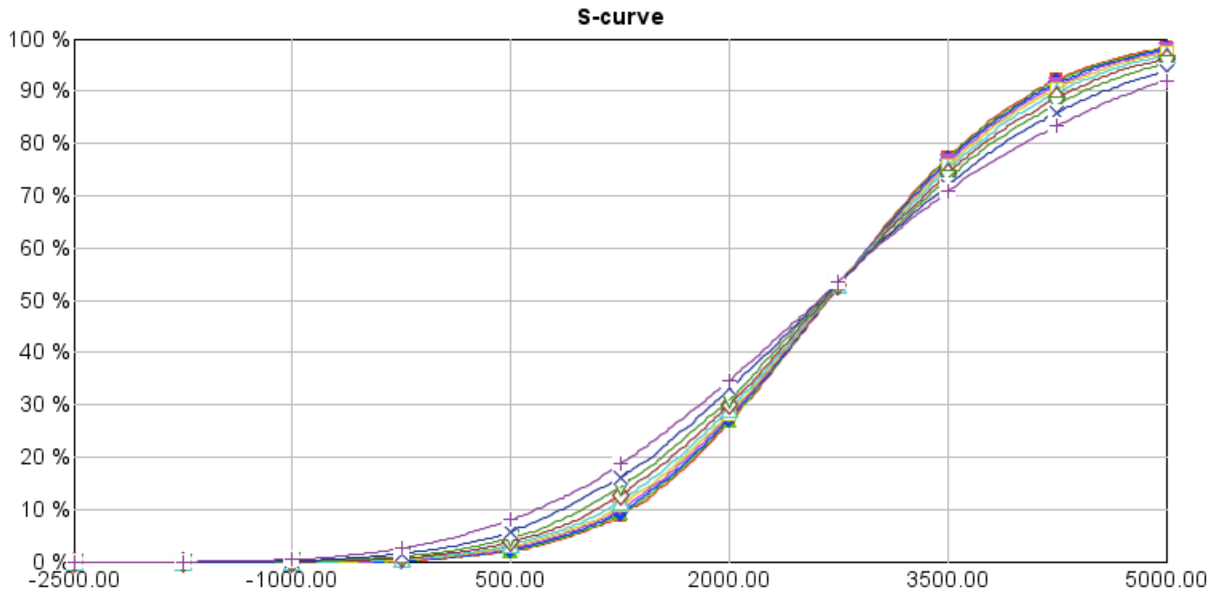


Figure 9. Density plot from Riscue software.

The probability that the net result of the project is less than 1 billion, as visualized in the S-curve plot above, is approximately 9%.

Problem - 1.j)

A sensitivity investigation of the results with respect to variations in the time correlation is implemented now. Here, there are 10 results according to the correlations 0,1, …, 1 respectively, as showed below:



	Mean	P10	P50	P90
Net result: 1	2679.05	1308.30	2669.09	4086.65
Net result: 2	2678.55	1300.18	2664.17	4103.80
Net result: 3	2677.99	1280.51	2659.52	4124.03
Net result: 4	2677.33	1246.97	2663.75	4159.47
Net result: 5	2676.53	1209.68	2658.38	4203.60
Net result: 6	2675.54	1159.12	2650.39	4251.95
Net result: 7	2674.31	1082.49	2639.90	4317.63
Net result: 8	2672.69	982.42	2635.09	4416.51
Net result: 9	2670.52	842.07	2613.56	4553.58
Net result: 10	2671.24	673.35	2613.28	4765.06

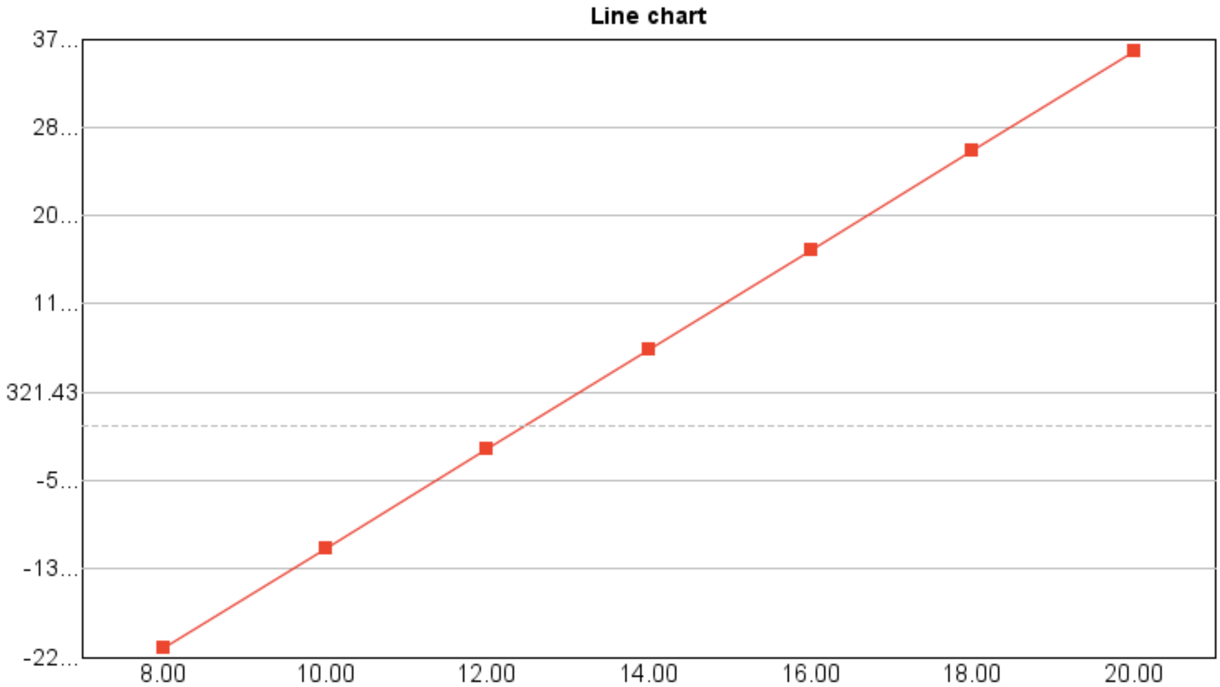
Figure 10. S-curve plot of different correlations parameters, from Riscue software

Firstly, as higher are the correlations as higher are the accumulation probabilities for the net result, for instance, the probability that the net result of the project is less than two billions is around 45% for correlation equals to 1 and 38% for correlation equals to 0,1. However, this is observed until net return equals to approximately 2,8 billions, where the accumulation probability is nearby 55% for all values of correlations. Finally, after 2,8 billions, it is possible to visualize that the lowest correlations start to lead the cumulative probability curve with higher probabilities against net returns. This means that the correlation is relevant between the consecutive early years of the project, but after some point, it turns to be not relevant anymore according to the sensitivity analyze.

Problem - 1.k)

Here, the study will determine how high the oil price needs to be to ensure a positive net result. In the Riscue software, the parameter oil price is switched to multivalued with prices 8,10, 12, …, 20 dollars per barrel of oil.

Again, the 10000 simulations of the net result were run and a line chart plot created:



	Mean	P10	P50	P90
Net result:1	-2146.52	-3471.26	-2144.76	-830.61
Net result:2	-1182.39	-2527.69	-1179.82	163.37
Net result:3	-218.26	-1594.41	-218.27	1159.47
Net result:4	745.87	-669.25	737.46	2171.05
Net result:5	1710.00	262.05	1696.52	3185.54
Net result:6	2674.13	1189.66	2655.62	4192.45
Net result:7	3638.27	2116.22	3611.67	5210.84

This plot shows that the price of the barrel of oil should be higher than approximately 12,5 dollars to the project be viable, or in other words, to get a positive expected net result.