

9th October, 2019

STK3405

Mandatory assignment 2 of 2

Submission deadline

Thursday 7th November 2019, 14:30 at Canvas (www.canvas.uio.no).

Instructions

You can choose between scanning handwritten notes or typing the solution directly on a computer (for instance with L^AT_EX). The assignment must be submitted as a single PDF file. Scanned pages must be clearly legible. The submission must contain your name, course and assignment number.

It is expected that you give a clear presentation with all necessary explanations. Remember to include all relevant plots and figures. Students who fail the assignment, but have made a genuine effort at solving the exercises, are given a second attempt at revising their answers. All aids, including collaboration, are allowed, but the submission must be written by you and reflect your understanding of the subject. If we doubt that you have understood the content you have handed in, we may request that you give an oral account.

In exercises where you are asked to write a computer program, you need to hand in the code along with the rest of the assignment. It is important that the submitted program contains a trial run, so that it is easy to see the result of the code.

Application for postponed delivery

If you need to apply for a postponement of the submission deadline due to illness or other reasons, you have to contact the Student Administration at the Department of Mathematics (e-mail: studieinfo@math.uio.no) well before the deadline.

All mandatory assignments in this course must be approved in the same semester, before you are allowed to take the final examination.

Complete guidelines about delivery of mandatory assignments:

uio.no/english/studies/admin/compulsory-activities/mn-math-mandatory.html

GOOD LUCK!

About the mandatory assignment for STK3405

In this assignment we shall use the simulation software RISCUETM. This can be downloaded for free at www.riscue.org. Choose your OS (Windows, Mac or Unix) by clicking the relevant icon on the left-hand side and follow the instructions on the webpage¹. In order to get your assignment approved, you must have at least 65 % correct answers.

Problem 1. A common probability distribution in risk analysis is the *lognormal distribution*. A stochastic variable, Y , is said to be lognormally distributed with parameters μ and σ if $X = \ln(Y)$ is normally distributed with expectation μ and standard deviation σ . We indicate this by writing $Y \sim \text{lognorm}(\mu, \sigma)$.

(a) Assume that $U \sim N(0, 1)$, i.e., U is standard normally distributed, and let $Y = \exp(\sigma U + \mu)$. Explain briefly why this implies that $Y \sim \text{lognorm}(\mu, \sigma)$.

(b) Assume that $Y \sim \text{lognorm}(\mu, \sigma)$, and let y_α be such that:

$$P(Y \leq y_\alpha) = \alpha\%, \text{ where } 0 \leq \alpha \leq 100.$$

Thus, according to this notation we e.g., have that y_{10} is the 10%-percentile in the $\text{lognorm}(\mu, \sigma)$ -distribution. Show that:

$$y_\alpha = \exp(\sigma u_\alpha + \mu),$$

where u_α is the $\alpha\%$ -percentile in the standard normal distribution.

(c) Show that:

$$\frac{y_{90}}{y_{50}} = \frac{y_{50}}{y_{10}}$$

(d) Assume that $y_{10} = 5.0$ and that $y_{90} = 20.0$. Find the corresponding value of y_{50} . Determine also the μ - and σ -values for this distribution. You may use that $u_{90} = 1.28155$.

[Solution: $\mu = 2.30259$, and $\sigma = 0.54087$.]

(e) Use RISCUETM to make a simulation model consisting of a single node. Name this node Y. (See Figure 1 below.) The node should have the following function:

$$\text{EXP}(\text{SUM}(\text{PRODUCT}(\text{NORMAL}()); 0.54087); 2.30259)) \quad (1)$$

Run a simulation on this model. Use e.g., 10000 iterations. Make an *S-curve* plot and a density plot for the resulting data node. Verify that the 10%-percentile and the 90%-percentile for the result node are (approximately) the desired value (i.e., respectively 5.0 and 20.0). Make a plot showing how these percentile values converge. Comment the results briefly.

¹Note that if you have a Mac with OS 10.8 (Mountain Lion) or newer, you have to switch off the protection preventing you from using software that has not been approved by Apple before you install RISCUETM. See detailed instruction on www.riscue.org.



Figure 1: Simulation model consisting of a single node.

We now turn to the problem of constructing a timeseries model where all the variables are lognormally distributed. In order to construct this model we start out with a series $\{W_n\}$ of independent standard normally distributed variables. Based on this series we then form a new series $\{U_n\}$ as follows:

$$U_1 = W_1, \quad U_n = a \cdot U_{n-1} + b \cdot W_n, \quad n = 2, 3, \dots$$

We observe that since U_{n-1} is included in the expression for U_n , $n = 2, 3, \dots$, the variables in this series become stochastically dependent.

(f) Determine the constants a and b such that $U_n \sim N(0, 1)$ and $\text{Corr}(U_n, U_{n-1}) = c$, $n = 2, 3, \dots$, where c is a suitable constant such that $|c| \leq 1$.

[Solution: $a = c$ and $b = \sqrt{1 - c^2}$.]

(g) We shall now use RISCUETM to make a simulation model as described above. Before you start creating the model, we recommend closing all open windows in RISCUETM. We then start out by creating an empty model by using the menu-command: *File>New*. You then get a blank window. Then we choose the menu-command *File>Save As...*, and name the model file *STK3405_mandatory2*.

In order to allow the correlation c to vary, we let this be represented in the model by a parameter. Such a parameter is created by first opening the *Parameter table window*, using the menu command: *Model>Parameter table...*. Then a parameter is created by using the menu command: *Model>Add Parameter*. Finally, we enter the relevant information into the columns of the parameter table ending up with a table like the one shown in Figure 2.

Name	ID	Def. value	Values	Multivalued	Info
Time correlation 0	0.5	0.1; 0.2; ... ; 1.0		<input checked="" type="checkbox"/>	Correlation between successive terms

Figure 2: Parameter table.

Tip: Pay special attention to what you enter into the *Values* column. As separator between the numbers make sure that you use a semicolon. Between 0.2 and 1.0 make sure that you type in exactly the following characters: *semicolon, space, period, period, period, space, semicolon, space*.

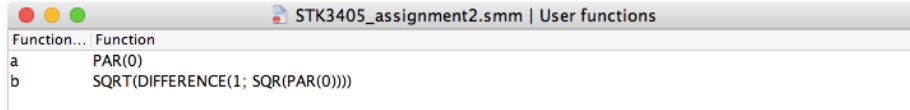
Tip: Note that the title field of the window containing the parameter table has the following name:

STK3405_mandatory2.smm | Parameter data

Thus, the first part of the title field indicates the name of the model file you are working on at the moment. If you are working on several different models at the same time, the menu-command *Model>Parameter table...* always opens the parameter table belonging to the frontmost model window. In such cases it may be a good idea to verify that you are editing the *correct* parameter table. This is done by checking the title field.

Note that the value in the *ID* column, i.e., 0, is generated automatically. This value is used to identify the parameter when it is used in formulas. We can then use the formula **PAR(0)**, where the argument is exactly this ID value. For the time being leave the checkbox in the *Multivalued* column unchecked. This indicates that the model will use the default parameter value given in the *Def. value*, i.e., 0.5 in the simulations.

In order to make the model easier to read, we also create some user defined functions. Thus, we open the *User function window* by using the *User>User Functions* menu command. We then create two such functions by using the *User>Add User Function* menu command, and enter the details as indicated in Figure 3.



Function...	Function
a	PAR(0)
b	SQRT(DIFFERENCE(1; SQR(PAR(0))))

Figure 3: User function table.

Tip: Note that the title field of the window containing the user defined functions has the following name:

STK3405_mandatory2.smm | User functions

Just as for the window containing the parameter table, the first part of the title field indicates the name of the model file you are working on at the moment. If you are working on several different models at the same time, it may be a good idea to verify that you are editing the *correct* user function table. This is again determined by the frontmost model window as you open the window containing the user function table.

We observe that **PAR(0)** refers to the correlation, i.e., c . Hence, the two user defined functions, called **a** and **b** respectively, are precisely the expressions derived for the constants a and b above.

We are now ready to create our timeseries model, a chain of $n = 12$ nodes as illustrated in Figure 4.

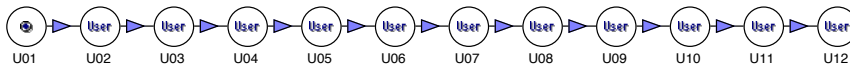


Figure 4: A time series model.

For the node **U01** we use the following formula:

$$\text{NORMAL} () \quad (2)$$

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, \dots, Y_{12} .

For all the other nodes we use the following formula:

$$\text{SUM}(\text{PRODUCT}(\mathbf{a}()); \text{INPUT}()); \text{PRODUCT}(\mathbf{b}()); \text{NORMAL}()) \quad (3)$$

Tip: In order to save time enter the formula (3) into node **U02**, and then copy this formula to the nodes **U03**, ..., **U12**. This is done as follows: Start out by entering formula (3) into **U02**. Then *right-click* the node **U02**, and choose *Copy formula* from the pop-up menu. Then select all the nodes **U03**, ..., **U12** by dragging a selection marquee around them (keep the mouse button pressed and move the cursor in order to create this selection marquee). While the nodes **U03**, ..., **U12** are selected, right-click on one of them (it does not matter which one). Then choose *Paste formula* from the pop-up menu.

Run a simulation on this model. Make a correlation table for the resulting data nodes by selecting the data nodes and using the menu command *Statistics>Tables>Correlations*, and verify that the correlation between any pair of consecutive nodes is approximately 0.5. Comment the results briefly.

(h) We now wish to transform the timeseries above into a new timeseries, $\{Y_n\}$ where all the variables are lognormally but not identically distributed. Since we know that $U_n \sim N(0, 1)$, $n = 1, 2, \dots, 12$, this can be achieved by letting:

$$Y_n = \exp(\sigma_n \cdot U_n + \mu_n), \quad n = 1, 2, \dots, 12.$$

where σ_n and μ_n are calculated based on the 10%- and 90%-percentiles of Y_n , $n = 1, 2, \dots, 12$, given in Table 1.

Calculate σ_n and μ_n , $n = 1, 2, \dots, 12$, based on the data in Table 1 using the result from (b).

Now, starting with the model in Figure 4, create a new row of nodes below the first row. Name the new nodes **Y01**, ..., **Y12** respectively. Then add arrows from **U01** to **Y01**, from **U02** to **Y02**, etc. until you end up with a diagram as the one shown in Figure 5.

For all the new nodes you should use a formula of the following form:

$$\text{EXP}(\text{SUM}(\text{PRODUCT}(\mathbf{X2}()); \text{INPUT}()); \mathbf{X1}()) \quad (4)$$

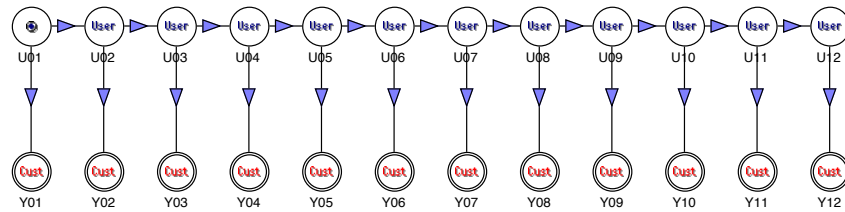
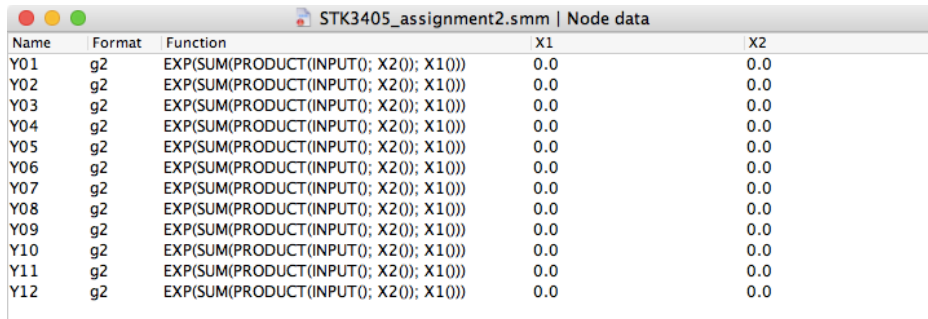


Figure 5: Two linked timeseries models.

where $X1()$ and $X2()$ are functions referring to the respective values of μ_n and σ_n , calculated based on the numbers in Table 1. The return values of these functions have to be entered into each of the nodes in the following way. Start by using the menu command *Model>Set Node Table Entries...*. As a result you should see the dialog box shown in Figure 6. Make sure that the checkboxes are checked as shown in the figure. In particular, make sure that $X1$ and $X2$ are checked.

Figure 6: The *Set Node Table Entries...* dialog box.

Click the **OK**-button to close the dialog box. Then select the nodes $Y01, \dots, Y12$, and choose the menu command *Model>Node Table*.



Name	Format	Function	X1	X2
Y01	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y02	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y03	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y04	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y05	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y06	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y07	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y08	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y09	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y10	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y11	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0
Y12	g2	EXP(SUM(PRODUCT(INPUT0; X20); X10))	0.0	0.0

Figure 7: Node table.

This brings up a table containing data for the nodes Y_{01}, \dots, Y_{12} . In the columns $X1$ and $X2$ you can enter the values of μ_n and σ_n for the 12 nodes.

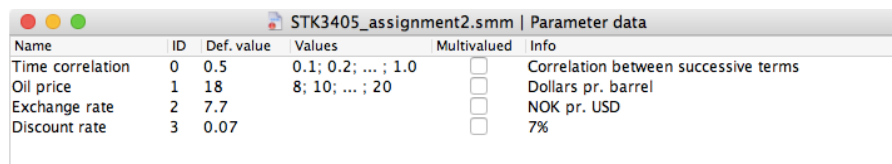
After you have entered all the values, run a simulation on the model. Make a correlation table for the resulting data nodes. Does it look like the correlation between any pair of consecutive nodes is approximately 0.5 in this case as well? Comment the results briefly.

Tip: In order to avoid having too many resulting data nodes, it is a good idea not to store results for the old nodes, i.e., U_{01}, \dots, U_{12} . This can be accomplished by selecting all these nodes simultaneously, right-click on one of them, and choose *Don't store results* from the pop-up menu.

(i) The timeseries $\{Y_n\}$ represents yearly production of oil (in millions of barrels) from an oilfield with "economic life" of 12 years. In order to calculate the income (in millions of Norwegian kroner) from this field in this period, we need to multiply the produced volumes by the oil price (in dollars per barrel) and the dollar exchange rate (in Norwegian kroner per dollar). Both the oil price and the exchange rate are of course uncertain quantities. However, it is common practice to enter these quantities into the model as parameters, and thus allowing us to run a sensitivity analysis.

In addition to this one typically *discounts* future incomes using a suitable discount rate. The result is called the *present value* of the incomes.

Create parameters for oil price, exchange rate and discount rate as shown in Figure 8.



Name	ID	Def. value	Values	Multivalued	Info
Time correlation	0	0.5	0.1; 0.2; ... ; 1.0	<input type="checkbox"/>	Correlation between successive terms
Oil price	1	18	8; 10; ... ; 20	<input type="checkbox"/>	Dollars pr. barrel
Exchange rate	2	7.7		<input type="checkbox"/>	NOK pr. USD
Discount rate	3	0.07		<input type="checkbox"/>	7%

Figure 8: Parameter table.

We then add three new nodes to the model: **Disc. income**, **Disc. cost** and **Net result**. From each of the nodes Y_{01}, \dots, Y_{12} we draw arrows down to

Disc. income. Finally, we draw arrows from Disc. income and Disc. cost to Net result. The resulting diagram should look like the one shown in Figure 9.

IMPORTANT! Create the arrow from Disc. income to Net result *before* you create the arrow from Disc. cost to Net result.

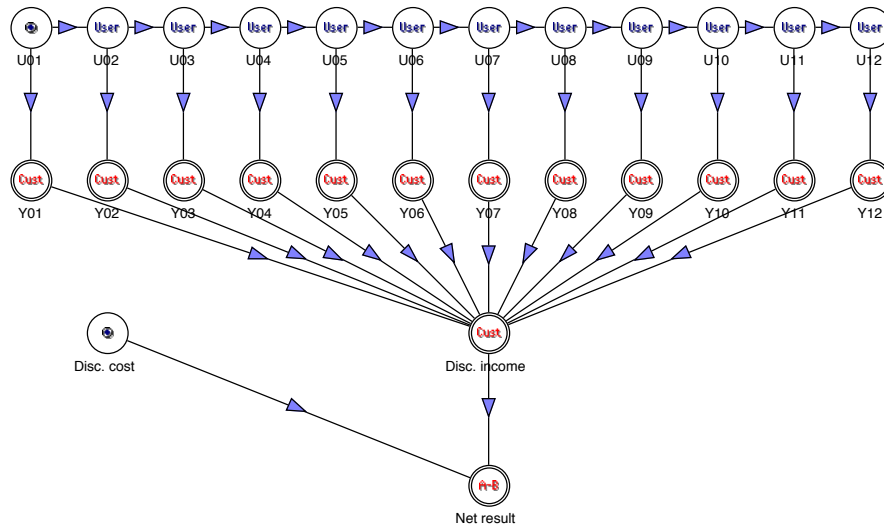


Figure 9: The complete model.

In the **Disc. income** node use the following formula:

$$\text{PRODUCT}(\text{INPUTNPV}(\text{PAR}(3)); \text{PAR}(1); \text{PAR}(2)), \quad (5)$$

while in the **Disc. cost** node enter the following formula:

$$\text{NORMAL}(6000; 1000). \quad (6)$$

Finally, enter the following formula into the **Net result** node:

$$\text{INPUTDIFF}(). \quad (7)$$

Make sure that the **Net result** node is the only node from which you store results during simulations. Give a brief description of how formula (5) works.

Run a simulation on this model. Create an *S-curve* plot and a *density* plot for the resulting data node (for the **Net result** node only). By using the S-curve estimate the probability that the net result of the project is less than 1 billion.

(j) We shall now investigate the sensitivity of the results with respect to variations in the time correlation, i.e., the parameter with ID = 0. Thus, we open the parameter window and switch the *Multivalued* option on for this parameter by clicking this checkbox.

Run a simulation with this setup. You should then end up with a window containing 10 data nodes, one for each value of the time correlation, ranging from 0.1 up to 1.0. Select all the data nodes and create a common *S-curve* plot. In what way does the correlation affect the distribution of the node **Net result**? Give a brief comment on this result.

(k) Finally, we wish to determine how high the oil price needs to be in order to ensure a positive net result. To do so, we switch the *Multivalued* option on for the oil price parameter, i.e., the parameter having ID = 1. Make sure that none of the other parameters have their *Multivalued* option switched on.

Run a simulation with this setup. You should then end up with a window containing 7 data nodes, one for each value of the oil price, i.e., for the values 8, 10, 12, ..., 20 (dollars per. barrel). Then use the menu command *Statistics>Set Chart Statistics...* You should then see a dialog box where you can select certain chart options. Make sure that the setup is exactly as shown in Figure 10 below.

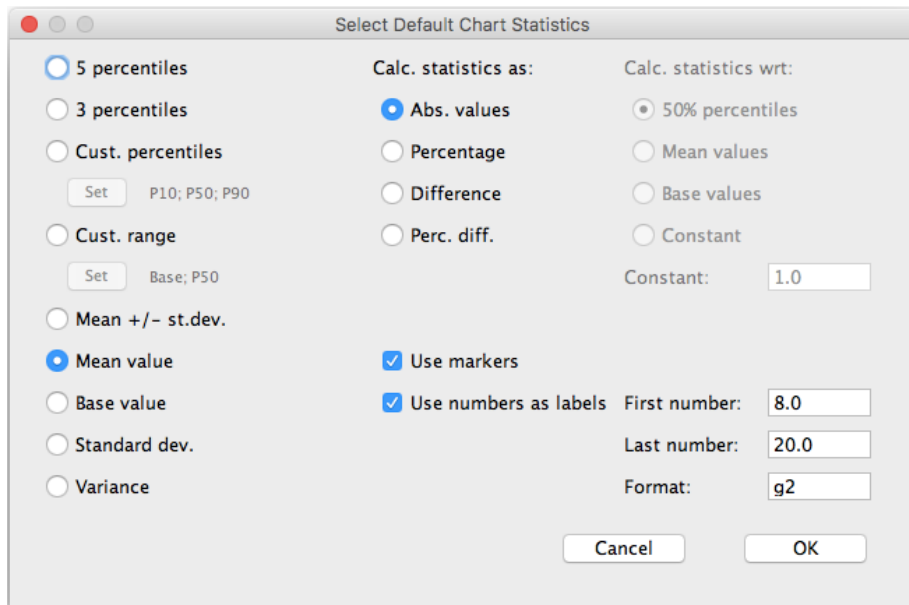


Figure 10: The *Set Chart Statistics...* dialog box.

Then select all the result nodes, and create a line chart. If you do everything right, the resulting plot should show the expected net result as a function of the oil price. From this plot you can read how high the oil price must be in order to get a positive expected net result. This is done by identifying the point where the curve intersects the zero-line (the horizontal gray dashed line in the plot). Comment this result.