

BOZ780 Assignment 1

DS de Gouveia
15003079

October 18, 2020

Contents

1	Question 1	2
1.1	Question 1 (a) - Stochastic program formulation	2
1.1.1	Assumptions	2
1.1.2	Problem formulation	2
1.2	Question 1 (b) - Chance constraint scenario approach	3
1.2.1	Problem formulation	3
1.3	Question 1 (c) - Credit shortfall	4
1.3.1	Assumptions	4
1.3.2	Stochastic problem formulation	4
1.3.3	Deterministic problem formulation	4
1.4	Code	5
2	Question 2 a	6
2.1	Model formulation	6
2.2	Histograms	7
3	Question 2 b	8
3.1	Model formulation	8
3.2	Results	9
3.3	Code snippets	9
4	Question 2 c	11
4.1	Model formulation	11
4.2	Results	12
4.3	Code snippets	12
5	Question 2 d	15
5.1	Model formulation	15
5.2	Results	17
6	Question 2 e	18
6.1	Model formulation	18
6.2	Code snippets	19
7	Question 2 f	21
7.1	Model formulation	21

Question 1

Question 1 (a) - Stochastic program formulation

Assumptions

- If a project is selected all of its cash flows will be implemented. If two or more projects are selected, their combined cashflow requirement is aggregated and compared to the available cash flow.
- The cashflows of each project may change but the NPV will remain the same, that is the NPV is independent on the cashflow.

Problem formulation

T set of years $T \in (1, 2, 3)$

I set of projects $I \in (1, \dots, 10)$

$a_t \triangleq$ cash in millions of dollars available for investment in year t , where $t \in T$

$c_{it} \triangleq$ cost of project i in millions of dollars for year t , where $t \in T, i \in I$

$n_i \triangleq$ net present value of project i , where $i \in I$

$x_i \triangleq \begin{cases} 1 & \text{project } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases} \quad i \in I, j \in J$

The objective of the investment is to increase profitability. The Net Present Value (NPV) is a means of assessing future profitability whilst accounting for inflation. Thus the objective of the model is to maximise the NPV of all projects whilst adhering to cash flow constraints.

$$\max z = \sum_{i \in I} x_i n_i \quad (1)$$

Subject to

$$\sum_{i \in I} x_i \tilde{c}_{it} \leq a_t \quad \forall t \in T \quad (2)$$

$$x_i \in \{0, 1\} \quad \forall i \in I \quad (3)$$

Of the two constraints, the first has been identified as an uncertain constraint with random variable \tilde{c}_{it} . Equation 2 states that for every year t the total cost of projects i must be less than or equal to the available cashflow a_t , noting that \tilde{c}_{it} is a random variable following a uniform distribution.

Question 1 (b) - Chance constraint scenario approach

The following question refers to the previously defined formulation in Question 1.

Problem formulation

The expected value formulation is disregarded and the uncertain constraints are reintroduced.

$$\max z = \sum_{i \in I} x_i n_i \quad (4)$$

Subject to

$$\sum_{i \in I} x_i \tilde{c}_{it} \leq a_t \quad \forall t \in T \quad (5)$$

$$x_i \in \{0, 1\} \quad \forall i \in I \quad (6)$$

Constraint 20 is applicable for 90% of the time. To accommodate this requirement, the constraint is expressed as

$$P\left(\sum_{i \in I} x_i \tilde{c}_{it} \leq a_t\right) \geq 0.9 \quad (7)$$

where P indicates the probability of this constraint being used. To solve this problem, a set of N constraints must be derived from a random sample of the uniformly distributed variable \tilde{c}_{it} which will convert equation 7 into the following form.

$$P(F(x, \zeta^v) \leq 0) \geq 1 - \alpha \quad v = 1, \dots, N \quad (8)$$

The scenario approach is implemented as follows.

$$P\left(\sum_{i \in I} x_i \tilde{c}_{it} \leq a_t\right) \geq 1 - 0.1 \quad (9)$$

From this form we can deduce that $\alpha = 0.1$ and $n = 10$. It is then assumed that a 99% confidence interval should be applied resulting in a $\delta = 0.01$. The confidence interval ensures that 99% of the samples will satisfies the constraint 90% of the time. These values are used to calculate the number of N samples using the following formula.

$$N = \lceil 2n\alpha^{-1} \ln\left(\frac{12}{\alpha}\right) + 2\alpha^{-1} \ln\left(\frac{2}{\delta}\right) + 2n \rceil \quad (10)$$

The known values are inserted into the formula and N is calculated.

$$N = \lceil 2(10)(0.1)^{-1} \ln\left(\frac{12}{0.1}\right) + 2(0.01)^{-1} \ln\left(\frac{2}{0.01}\right) + 2(10) \rceil \quad (11)$$

$$N = 1083.465 \approx 1084 \quad (12)$$

The constraint may be reformulated as

$$\sum_{i \in I} x_i c_{it}^q \leq a_t \quad \forall q \in Q, \forall t \in T \quad (13)$$

where $|Q| = N = 1084$ is the sample set.

Question 1 (c) - Credit shortfall

Assumptions

The following question refers to the previously defined formulation in Question 1.

Stochastic problem formulation

The non-deterministic formulation shall be re-formulated into a recourse programme. Doing so will include a recourse variable in the constraint that captures the deficient cashflow amount which is the difference between a_t and the total cost of the projects. The deficiency is then penalised in the objective function using the 15% interest rate as a scalar quantity. The following recourse programme extends the formulation in Question 1.

$$\begin{aligned} e_t &\triangleq \text{compounded interest rate applied in year } t, \text{ where } t \in T \text{ for } \{1.15, 1.32, 1.52\} \\ \delta_t^- &\triangleq \text{cashflow deficit in millions of dollars for year } t, \text{ where } t \in T \\ \delta_t^+ &\triangleq \text{cashflow surplus in millions of dollars for year } t, \text{ where } t \in T \end{aligned}$$

The δ_t^- recourse variable captured the cashflow deficient. Including δ_t^+ is merely a formality to accommodate the possibility of surplus cashflow, which could only occur if the constraints are tightened (reduced number of projects). For this reason, δ_t^+ is not penalised ergo, included in the objective function.

$$\max z = \sum_{i \in I} x_i n_i - \sum_{t \in T} E_{\tilde{c}_{it}} [e_t \delta_t^- (\tilde{c}_{it})] \quad (14)$$

Subject to

$$\sum_{i \in I} x_i \tilde{c}_{it} + \delta_t^+ (\tilde{c}_{it}) - \delta_t^- (\tilde{c}_{it}) = a_t \quad \forall t \in T \quad (15)$$

$$x_i \in \{0, 1\} \quad \forall i \in I \quad (16)$$

$$\delta_t^+ (\tilde{c}_{it}), \delta_t^- (\tilde{c}_{it}) \geq 0 \quad \forall i \in I, \forall t \in T \quad (17)$$

Deterministic problem formulation

To formulate the stochastic problem into a deterministic equivalent, discrete realisations of the random variable must be created.

$$\tilde{c}_{it} \approx \{(p_1^{r_1}, y_1^{r_1, r_2})\}_{r_1 \in R_1} \quad (18)$$

where $p_1^{r_1}$ is the probability of impact $y_1^{r_1}$ occurring for realisation $r_1 \in R_1$ where R_1 is the set of bins chosen. In this instance, five bins have been selected. Using the R code snippet 1.4 the probability and impact for each bin are calculated and results are summarised in Table 1. For each project i the same impact is applicable as they are derived from the same random variable.

The deterministic realisations of the random variable are then included in the model. For each realisation a constraint is added to ensure that under those condition the model remains feasible. The probability is then used in the objective function to penalise the recourse variable proportionally to the likelihood of the particular realisation of occurring.

Table 1: Probability and impact for each bin.

Bin (r_1)	Probability ($p_1^{r_1}$)	Impact ($y_1^{r_1}$)
1	0.19817	5.52
2	0.19946	5.76
3	0.20177	6.00
4	0.20019	6.24
5	0.20041	6.48

The deterministic equivalent can be expressed as

$$\max z = \sum_{i \in I} x_i n_i - \sum_{t \in T} \sum_{r_1=1}^5 p_1^{r_1} (e_t \delta_{t,r_1}^-) \quad (19)$$

Subject to

$$\sum_{i \in I} x_i y_1^{r_1} + \delta_{t,r_1}^+ - \delta_{t,r_1}^- = a_t \quad \forall t \in T, \forall r_1 \in R_1 \quad (20)$$

$$x_i \in \{0, 1\} \quad \forall i \in I \quad (21)$$

$$\delta_t^+, \delta_t^- \geq 0 \quad \forall t \in T \quad (22)$$

Code

Calculation for the the probability and impact per bin.

```
set.seed(20201010)
```

```
fn.UniDiscrete <- function(start, end, bins){
  hist <- hist(x = runif(n=100000, min=start, max=end),
               breaks = seq(start,end,(end-start)/bins))
  Probability <- hist$counts/100000
  Impact <- hist$mid
  solution <- data.frame(Probability,Impact)
  return(solution)
}
```

```
fn.UniDiscrete(start = 5.4, end = 6.6, bins = 5)
```

Question 2 a

Model formulation

The following stochastic formulation is derived as a binary integer programme with stochastic constraints.

$$\begin{aligned} S^n & \text{ set of normal regions } S^n \in (1, 2, 3) \\ S^t & \text{ set of tumour regions } S^t \in (1, 2, 3) \\ I & \text{ set of beams } I \in (1, \dots, 6) \end{aligned}$$

$$x_{is^n} \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used on region } s^n, \text{ where } i \in I, s^n \in S^n \\ 0 & \text{otherwise} \end{cases}$$

$$y_i \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used, where } i \in I \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} r_{is^n} & \triangleq \text{ radiation from beam } i \text{ hitting normal region } s^n, \text{ where } s^n \in S^n \\ r_{is^t} & \triangleq \text{ radiation from beam } i \text{ hitting tumour region } s^t, \text{ where } s^t \in S^t \end{aligned}$$

The objective function is formulated to maximise the radiation hitting tumour regions r_{s^t} . It should be noted that normal regions and tumour regions share a similar set S but refer specific variations as S^n and S^t . The fifth and fourth beams are denoted as random variables such that $\tilde{r}_{3s^n} = \{\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3\}$ and $\tilde{r}_{4s^n} = \{\tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3\}$.

$$\max z = \sum_{i \in I} \sum_{s^t \in S^t} x_{is^t} r_{is^t} \quad (23)$$

Subject to

$$\sum_{i=\{1,2,5,6\}} x_{i1} r_{i1} + x_{31} \tilde{\zeta}_1 + x_{41} \tilde{\eta}_1 \leq 40 \quad (24)$$

$$\sum_{i=\{1,2,5,6\}} x_{i2} r_{i2} + x_{32} \tilde{\zeta}_2 + x_{42} \tilde{\eta}_2 \leq 40 \quad (25)$$

$$\sum_{i=\{1,2,5,6\}} x_{i3} r_{i3} + x_{33} \tilde{\zeta}_3 + x_{43} \tilde{\eta}_3 \leq 40 \quad (26)$$

$$\sum_{s^n \in S^n} x_{is^n} = 3y_i \quad \forall i \in I \quad (27)$$

$$x_{is^n}, y_i \geq 0 \quad \forall i \in I, \forall s^n \in S^n \quad (28)$$

Constraint 24,25 and 26 ensure that the radiation hitting a normal region is below the hazardous threshold. Nothing that these constraints are formulated using random variables. Constraint 27 ensures that if a beam is used on one region, it shall remain active for all other regions. Constraint 28 is the non-negativity constraint.

Histograms

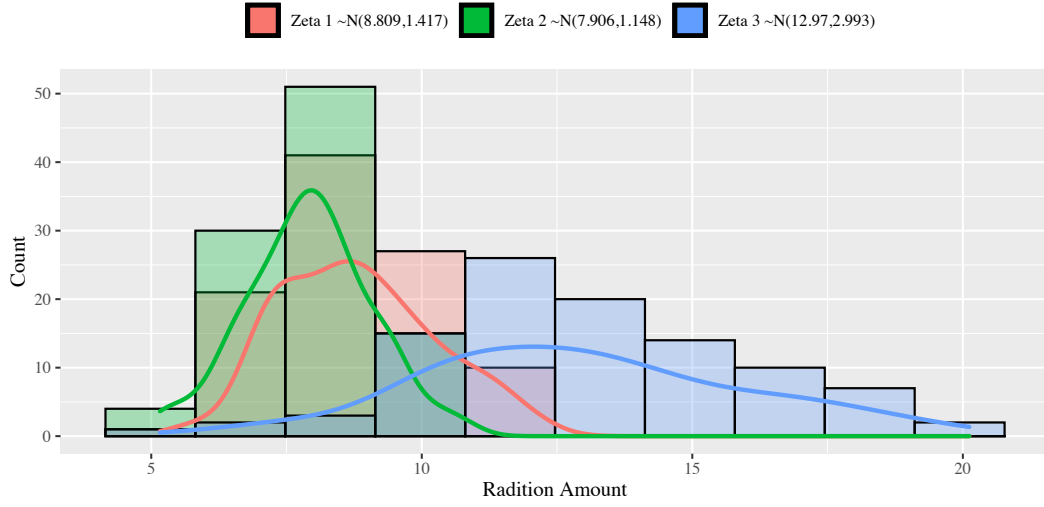


Figure 1: Distribution of beam 3 radiation on normal region for random variable $\zeta_1, \zeta_2, \zeta_3$. The smooth line is a dimensionless expression of the random variables spread (density plot) whereas the histogram illustrates the number of occurrences per bin.

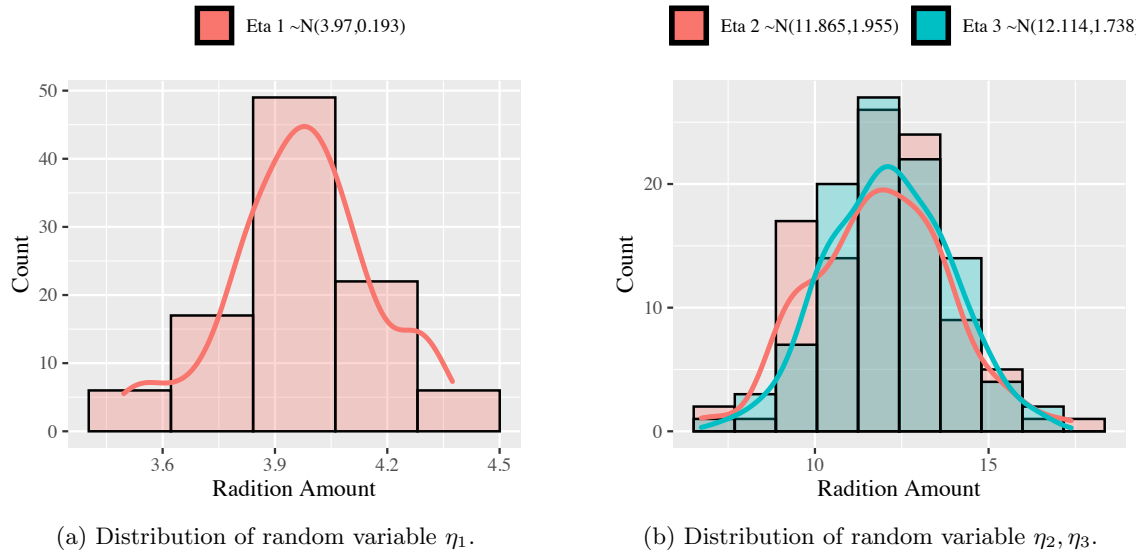


Figure 2: Distribution of beam 4 radiation on normal region for random variable η_1, η_2, η_3 . The smooth line is a dimensionless expression of the random variables spread (density plot) whereas the histogram illustrates the number of occurrences per bin. These plots were expressed separately as the magnitude of the counts differed greatly.

Question 2 b

Model formulation

S^n set of normal regions $S^n \in (1, 2, 3)$

S^t set of tumour regions $S^t \in (1, 2, 3)$

I set of beams $I \in (1, \dots, 6)$

$$x_{is^n} \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used on region } s^n, \text{ where } i \in I, s^n \in S^n \\ 0 & \text{otherwise} \end{cases}$$

$$y_i \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used, where } i \in I \\ 0 & \text{otherwise} \end{cases}$$

$r_{is^n} \triangleq$ radiation from beam i hitting normal region s^n , where $s^n \in S^n$

$r_{is^t} \triangleq$ radiation from beam i hitting tumour region s^t , where $s^t \in S^t$

The objective function maximises radiation hitting tumour regions r_{s^t} .

$$\max z = \sum_{i \in I} \sum_{s^t \in S^t} x_{is^t} r_{is^t} \quad (29)$$

The fifth and fourth beams are denoted as random variables such that $\tilde{r}_{3s^n} = \{\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3\}$ and $\tilde{r}_{4s^n} = \{\tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3\}$. A deterministic equivalent is formulated using the expected value of the random variables. These values are estimated using the mean from the provided samples resulting in the following estimated parameters.

$$\tilde{r}_{3s^n} = \{8.809, 7.906, 12.97\} \quad (30)$$

$$\tilde{r}_{4s^n} = \{3.97, 11.865, 12.114\} \quad (31)$$

Subject to

$$\sum_{i=\{1,2,5,6\}} x_{i1} r_{i1} + x_{31} E(\tilde{\zeta}_1) + x_{41} E(\tilde{\eta}_1) \leq 40 \quad (32)$$

$$\sum_{i=\{1,2,5,6\}} x_{i2} r_{i2} + x_{32} E(\tilde{\zeta}_2) + x_{42} E(\tilde{\eta}_2) \leq 40 \quad (33)$$

$$\sum_{i=\{1,2,5,6\}} x_{i3} r_{i3} + x_{33} E(\tilde{\zeta}_3) + x_{43} E(\tilde{\eta}_3) \leq 40 \quad (34)$$

$$\sum_{s^n \in S^n} x_{is^n} = 3y_i \quad \forall i \in I \quad (35)$$

$$x_{is^n}, y_i \geq 0 \quad \forall i \in I, \forall s^n \in S^n \quad (36)$$

Constraint 24,25 and 26 ensure that the radiation hitting a normal region is below the hazardous threshold, estimated as $E(\tilde{\eta}) = \{8.81, 7.91, 12.97\}$ and $E(\tilde{\zeta}) = \{3.97, 11.86, 12.11\}$. Constraint 27 ensures that if a beam is used on one region, it shall remain active for all other regions. Constraint 28 is the non-negativity constraint.

Results

Using R and the LpSolve package, the code snippet 1.4 yields the resulting solution.

$$\max z = 151 \quad (37)$$

$$x_{is^n} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (38)$$

Beams 2,3,4 and 6 are to be used to maximise the objective function. The radiation hitting each normal region is $r_{s^n} = \{32.78, 36.77, 38.08\}$ respectively, and the radiation hitting each tumour region is $r_{s^t} = \{47, 53, 51\}$ respectively. If a beam is used on one region it is used on all regions, as per constraint, thus all constraints are adhered to.

Code snippets

```
#solve expected value model
```

```
library(tidyverse)
```

```
library(lpSolve)
```

```
matMaker <- function(ones,len){
  temp <- rep(0,len)
  temp[ones] <- 1
  return(temp)
}
```

```
matMaker(c(1,2,3),100)
```

```
data.meanRadition <- read.csv(file = "RadiationData.csv", sep = ";")
%>% summarise_all("mean")
%>% as.numeric()
```

```
#           N N N T T T
data.Radiation <- c(16,12,8,20,12,6,
                  12,10,6,18,15,8,
                  data.meanRadition[1:3],13,10,17,
                  data.meanRadition[4:6],6,18,16,
                  9,4,11,13,5,14,
                  8,7,7,10,10,10)
```

```
data.Radiation.Normal <- c(16,12,8,
                          12,10,6,
                          data.meanRadition[1:3],
                          data.meanRadition[4:6],
                          9,4,11,
                          8,7,7)
```

```
data.Radiation.tumour <- c(20,12,6,
```

```

18,15,8,
13,10,17,
6,18,16,
13,5,14,
10,10,10)

# Define lp solve parameters

obj <- c(data.Radiation.tumour,rep(0,6))

constraint1.value <- c(matMaker(c(1,4,7,10,13,16),18)*data.Radiation.Normal,rep(0,6))
constraint2.value <- c(matMaker(c(2,5,8,11,14,17),18)*data.Radiation.Normal,rep(0,6))
constraint3.value <- c(matMaker(c(3,6,9,12,15,18),18)*data.Radiation.Normal,rep(0,6))

constraint4.value <- matrix(c(matMaker(c(1:3,19),24)+matMaker(19,24)*-4,
                             matMaker(c(4:6,20),24)+matMaker(20,24)*-4,
                             matMaker(c(7:9,21),24)+matMaker(21,24)*-4,
                             matMaker(c(10:12,22),24)+matMaker(22,24)*-4,
                             matMaker(c(13:15,23),24)+matMaker(23,24)*-4,
                             matMaker(c(16:18,24),24)+matMaker(24,24)*-4),
                             ncol = 24,byrow = T)

constraint.dir <- c(rep("<=",3),
                    rep("=",6))
constraint.rhs <- c(rep(40,3),
                    rep(0,6))

q2b.lp <- lp(direction = "max",
             objective.in = obj,
             const.mat = rbind(constraint1.value,
                               constraint2.value,
                               constraint3.value,
                               constraint4.value),
             const.dir = constraint.dir,
             const.rhs = constraint.rhs,
             all.bin = T)

q2b.lp
beams <- q2b.lp$solution[1:18] %>% matrix(ncol = 3,byrow = T)
q2b.lp$solution[19:24] %>% matrix(nrow = 6,byrow = T)

as.matrix(beams *matrix(data.Radiation.Normal,ncol = 3,byrow = T))%>%
  colSums() %>%
  round(digits = 2)
as.matrix(beams *matrix(data.Radiation.tumour,ncol = 3,byrow = T))%>% colSums()

```

Question 2 c

Model formulation

S^n set of normal regions $S^n \in (1, 2, 3)$

S^t set of tumour regions $S^t \in (1, 2, 3)$

I set of beams $I \in (1, \dots, 6)$

$$x_{is^n} \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used on region } s^n, \text{ where } i \in I, s^n \in S^n \\ 0 & \text{otherwise} \end{cases}$$

$$y_i \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used, where } i \in I \\ 0 & \text{otherwise} \end{cases}$$

$r_{is^n} \triangleq$ radiation from beam i hitting normal region s^n , where $s^n \in S^n$

$r_{is^t} \triangleq$ radiation from beam i hitting tumour region s^t , where $s^t \in S^t$

The fifth and fourth beams are denoted as random variables such that $\tilde{r}_{3s^n} = \{\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3\}$ and $\tilde{r}_{4s^n} = \{\tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3\}$. A deterministic equivalent can be formulated using chance constraint programming and the scenario approach by generating N scenarios sampled from each random variable to ensure a set of N constraints are satisfied 99% of the time.

The constraints containing random variables are expressed as

$$P\left(\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + x_3\tilde{\zeta}_1 + x_4\tilde{\eta}_1 \leq 40\right) \geq 1 - 0.01 \quad (39)$$

$$P\left(\sum_{i=\{1,2,5,6\}} x_{i2}r_{i2} + x_3\tilde{\zeta}_2 + x_4\tilde{\eta}_2 \leq 40\right) \geq 1 - 0.01 \quad (40)$$

$$P\left(\sum_{i=\{1,2,5,6\}} x_{i3}r_{i3} + x_3\tilde{\zeta}_3 + x_4\tilde{\eta}_3 \leq 40\right) \geq 1 - 0.01 \quad (41)$$

where P is the probability of the constraint being satisfied 99% of the time. The percentage is expressed as $1 - 0.01 = 0.99$ such that $\alpha = 0.01$. There are 2 random variables per constraint resulting in $n = 2$. A 99% confidence interval is selected, setting $\delta = 0.01$. The confidence interval ensures that the sample N is large enough for the constraint to be adhered to. N is calculated as follows.

$$N = \lceil 2n\alpha^{-1} \ln\left(\frac{12}{\alpha}\right) + 2\alpha^{-1} \ln\left(\frac{2}{\delta}\right) + 2n \rceil \quad (42)$$

$$N = \lceil 2(2)(0.01)^{-1} \ln\left(\frac{12}{(0.01)}\right) + 2(0.01)^{-1} \ln\left(\frac{2}{(0.01)}\right) + 2(2) \rceil \quad (43)$$

$$N = 3899.694 \approx 3900 \quad (44)$$

This implies $\{\zeta_1^q, \eta_1^q\}$, $\{\zeta_2^q, \eta_2^q\}$ and $\{\zeta_3^q, \eta_3^q\}$ will be sampled for $q \in N$ instances to generated N constraints. The deterministic formulation is defines as follows.

$$\max z = \sum_{i \in I} \sum_{s^t \in S^t} x_{is^t} r_{is^t} \quad (45)$$

Subject to

$$\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + x_{31}\zeta_1^q + x_{41}\eta_1^q \leq 40 \quad \forall q \in N \quad (46)$$

$$\sum_{i=\{1,2,5,6\}} x_{i2}r_{i2} + x_{32}\zeta_2^q + x_{42}\eta_2^q \leq 40 \quad \forall q \in N \quad (47)$$

$$\sum_{i=\{1,2,5,6\}} x_{i3}r_{i3} + x_{33}\zeta_3^q + x_{43}\eta_3^q \leq 40 \quad \forall q \in N \quad (48)$$

$$\sum_{s^n \in S^n} x_{is^n} = 3y_i \quad \forall i \in I \quad (49)$$

$$y_i \geq 0 \quad \forall i \in I \quad (50)$$

$$x_{is^n} \geq 0 \quad \forall i \in I, \forall s^n \in S^n \quad (51)$$

Constraints 46, 47 and 48 ensure that the radiation hitting a normal region is below the hazardous threshold, estimated by generating $N = 3900$ constraints. Thus these constraints are a condensed expression of all possible constraints. The constraint values are sampled from a distribution to ensure a 99% confidence interval. The remaining constraints are defined as in the previous section.

Results

To implement the solution, a 18 by 3904 matrix was formulated in which columns 7 to 12 denote the random variables. Each row in columns 7 to 12 is a value sampled from a normal distribution using the `rnorm()` function. The input parameters are the respective means and standard deviations of the random variables. Using R and the `LpSolve` package, the code snippet 4.2 yields the resulting solution.

$$\max z = 41 \quad (52)$$

$$x_{is^n} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (53)$$

Only beam 2 is be used to maximise the objective function when compared to the expected value approach which activated beams 2,3,4 and 6. The radiation hitting the each tumour region is $r_{st} = \{18, 15, 8\}$ respectively. The result is far lower than the expected value approach, indicating a far more restricting constraint. This approximation is more conservative yet yields a lower objective function and r_{st} . No constraint are violated.

Code snippets

```
#solve approximation model
```

```
library(tidyverse)
```

```
library(lpSolve)
```

```

matMaker <- function(ones,len){
  temp <- rep(0,len)
  temp[ones] <- 1
  return(temp)
}

data.meanRadition <- read.csv(file = "RadiationData.csv", sep = ";")
%>% summarise_all("mean")
%>% as.numeric()
data.sdRadition <- read.csv(file = "RadiationData.csv", sep = ";")
%>% summarise_all("sd")
%>% as.numeric()

N = 3900

data.Radiation.Normal <- c(16,12,8,
                           12,10,6,
                           0,0,0,
                           0,0,0,
                           9,4,11,
                           8,7,7)

data.Radiation.tumour <- c(20,12,6,
                           18,15,8,
                           13,10,17,
                           6,18,16,
                           13,5,14,
                           10,10,10)

constraint1.value <- rep(data.Radiation.Normal,N) %>% matrix(ncol = 18,byrow = T)

for (i in 7:12){
  constraint1.value[,i] <- round(rnorm(n = N,
    mean =data.meanRadition[i-6] ,
    sd = data.sdRadition[i-6]),
    digits = 4)
}
constraint1.value <- cbind(constraint1.value,matrix(data = 0,nrow = N,ncol = 6))

# Define lp solve parameters

obj <- c(data.Radiation.tumour,rep(0,6))

constraint2.value <- matrix(c(matMaker(c(1:3,19),24)+matMaker(19,24)*-4,
                             matMaker(c(4:6,20),24)+matMaker(20,24)*-4,
                             matMaker(c(7:9,21),24)+matMaker(21,24)*-4,

```

```

matMaker(c(10:12,22),24)+matMaker(22,24)*-4,
matMaker(c(13:15,23),24)+matMaker(23,24)*-4,
matMaker(c(16:18,24),24)+matMaker(24,24)*-4),
ncol = 24,byrow = T)

constraint.dir <- c(rep("<=",N),
                    rep("=",6))
constraint.rhs <- c(rep(40,N),
                    rep(0,6))

q2c.lp <- lp(direction = "max",
             objective.in = obj,
             const.mat = rbind(constraint1.value,constraint2.value),
             const.dir = constraint.dir,
             const.rhs = constraint.rhs,
             all.bin = T)

q2c.lp
beams <- q2b.lp$solution[1:18] %>% matrix(ncol = 3,byrow = T)
q2b.lp$solution[19:24] %>% matrix(nrow = 6,byrow = T)
beams

```

Question 2 d

Model formulation

S^n set of normal regions $S^n \in (1, 2, 3)$

S^t set of tumour regions $S^t \in (1, 2, 3)$

I set of beams $I \in (1, \dots, 6)$

$$x_{is^n} \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used on region } s^n, \text{ where } i \in I, s^n \in S^n \\ 0 & \text{otherwise} \end{cases}$$

$$y_i \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used, where } i \in I \\ 0 & \text{otherwise} \end{cases}$$

$r_{is^n} \triangleq$ radiation from beam i hitting normal region s^n , where $s^n \in S^n$

$r_{is^t} \triangleq$ radiation from beam i hitting tumour region s^t , where $s^t \in S^t$

The fifth and fourth beams are denoted as random variables such that $\tilde{r}_{3s^n} = \{\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3\}$ and $x_4\tilde{r}_{4s^n} = \{\tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3\}$. Figure 1 and Figure 2 provide justification to assume that the random variables are rotationally invariant (symmetrical distributions) thus permitting the use of chance constraint approximation methods. The constraints may be modelled as follows.

$$P\left(\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + x_3\tilde{\zeta}_1 + x_4\tilde{\eta}_1 \leq 40\right) \geq 1 - 0.01 \quad (54)$$

$$P\left(\sum_{i=\{1,2,5,6\}} x_{i2}r_{i2} + x_3\tilde{\zeta}_2 + x_4\tilde{\eta}_2 \leq 40\right) \geq 1 - 0.01 \quad (55)$$

$$P\left(\sum_{i=\{1,2,5,6\}} x_{i3}r_{i3} + x_3\tilde{\zeta}_3 + x_4\tilde{\eta}_3 \leq 40\right) \geq 1 - 0.01 \quad (56)$$

There are two random variables per constraint resulting in $m = 2$ and $\alpha = 0.01$ which allows Ω to be calculated as follows.

$$\Omega = \left(\sqrt{2 \ln \frac{2}{0.01}}\right) \quad (57)$$

$$= 3.255247 \approx 3.255 \quad (58)$$

Using Ω and converting the equations in the approximation method's standard form, It should be noted that x_i is a binary variable and as such there is no impact of raising x_i to any power. This property assists in reducing equation 54 as follows.

Table 2: Fixed and variable components used in the approximation approach.

s^n	$\bar{\zeta}_{s^n}$	$\tilde{\zeta}'_{s^n}$	$\bar{\eta}_{s^n}$	$\tilde{\eta}'_{s^n}$	$\zeta_{s^n}^2$	$\eta_{s^n}^2$
1	8.809	N(0,1.417)	3.970	N(0,0.193)	2.008	0.037
2	7.906	N(0,1.148)	11.865	N(0,1.955)	1.319	3.824
3	12.97	N(0,2.993)	12.114	N(0,1.738)	8.959	3.021

$$\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} - 40 + x_{31}(\bar{\zeta}_1 + \tilde{\zeta}'_1) + x_{41}(\bar{\eta}_1 + \tilde{\eta}'_1) \leq 0 \quad (59)$$

$$\left[\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + x_{31}\bar{\zeta}_1 + x_{41}\bar{\eta}_1 - 40 \right] + \Omega((x_{31}((\tilde{\zeta}'_1)^2 \times 1^2)^{\frac{1}{2}} + x_{41}((\tilde{\eta}'_1)^2 \times 1^2)^{\frac{1}{2}}) \leq 0 \quad (60)$$

$$\left[\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + 8.809x_{31} + 3.970x_{41} - 40 \right] + 3.255(1.417x_{31} + 0.193x_{41}) \leq 0 \quad (61)$$

$$\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + 13.421x_{31} + 4.598x_{41} \leq 40 \quad (62)$$

The random variables are approximated as $\tilde{\zeta}_1 \approx 13.421$ and $\tilde{\eta}_1 \approx 4.598$. Equation 55 and Equation 56 were reduced using the same method, resulting in $\tilde{\zeta}_2 \approx 11.644$, $\tilde{\eta}_2 \approx 18.230$, $\tilde{\zeta}_3 \approx 22.714$ and $\tilde{\eta}_3 \approx 17.772$. Leading to the following deterministic formulation.

$$\max z = \sum_{i \in I} \sum_{s^t \in S^t} x_{is^t}r_{is^t} \quad (63)$$

Subject to

$$\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + 13.421x_{31} + 4.598x_{41} \leq 40 \quad (64)$$

$$\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + 11.644x_{32} + 18.230x_{42} \leq 40 \quad (65)$$

$$\sum_{i=\{1,2,5,6\}} x_{i1}r_{i1} + 22.714x_{33} + 17.772x_{43} \leq 40 \quad (66)$$

$$\sum_{s^n \in S^n} x_{is^n} = 3y_i \quad \forall i \in I \quad (67)$$

$$y_i \geq 0 \quad \forall i \in I \quad (68)$$

$$x_{is^n} \geq 0 \quad \forall i \in I, \forall s^n \in S^n \quad (69)$$

Constraints 64, 65 and 66 ensure that the radiation hitting a normal region is below the hazardous threshold, estimated using the analytical approximation approach. The remaining constraints are defined as in the previous section.

Results

To implement the solution, a 18 by 3904 matrix was formulated in which columns 7 to 12 denote the random variables. Each row in columns 7 to 12 is a value sampled from a normal distribution using the `rnorm()` function. The input parameters are the respective means and standard deviations of the random variables. Using R and the `LpSolve` package, the code snippet 4.2 yields the resulting solution.

$$\max z = 113 \tag{70}$$

$$x_{is^n} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{71}$$

Using the scenario approach resulted in only beam 2 being selected whereas the analytical approximation includes beam 4 and 5. Compared to the expected value approach, one less beam is used and beam 5 is activated instead of beam 6 resulting in a marginally lower object function value. The radiation hitting each tumour region is $r_{st} = \{37, 38, 38\}$ respectively and the radiation hitting each normal region is $r_{st} = \{25.60, 32.23, 34.77\}$ respectively. This approximation slightly more conservative than the expected value approach yet far more lenient than the scenario approach. No constraints are violated.

Question 2 e

Model formulation

$$x_i \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used, where } i \in I \text{ and } I = \{1, \dots, 6\} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} r_{is^n} &\triangleq \text{radiation from beam } i \text{ hitting normal region } s^n, \text{ where } s^n \in S^n \text{ and } S^n = \{1, 2, 3\} \\ r_{is^t} &\triangleq \text{radiation from beam } i \text{ hitting tumour region } s^t, \text{ where } s^t \in S^t \text{ and } S^t = \{1, 2, 3\} \end{aligned}$$

The fifth and fourth beams are denoted as random variables such that $\tilde{r}_{3s^n} = \{\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3\}$ and $x_4\tilde{r}_{4s^n} = \{\tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3\}$. Two additional variables are introduced to initiate the recourse model formulation.

$$\begin{aligned} \delta_{s^n}^- &\triangleq \text{excess levels of radiation hitting normal region } s^n, \text{ where } s^n \in S^n. \\ \delta_{s^n}^+ &\triangleq \text{reduced levels of radiation hitting normal region } s^n, \text{ where } s^n \in S^n. \end{aligned}$$

$$\max z = \sum_{i \in I} \sum_{s^t \in S^t} x_i r_{is^t} - \sum_{s^n \in S^n} E_{\tilde{\zeta}, \tilde{\eta}}(2\delta_{s^n}^-(\tilde{\zeta}, \tilde{\eta})) \quad (72)$$

Subject to

$$\sum_{i \in \{1, 2, 5, 6\}} x_i r_{i1} + x_3 \tilde{\zeta}_1 + x_4 \tilde{\eta}_1 + \delta_1^+(\tilde{\zeta}, \tilde{\eta}) - \delta_1^-(\tilde{\zeta}, \tilde{\eta}) = 40 \quad (73)$$

$$\sum_{i \in \{1, 2, 5, 6\}} x_i r_{i2} + x_3 \tilde{\zeta}_2 + x_4 \tilde{\eta}_2 + \delta_2^+(\tilde{\zeta}, \tilde{\eta}) - \delta_2^-(\tilde{\zeta}, \tilde{\eta}) = 40 \quad (74)$$

$$\sum_{i \in \{1, 2, 5, 6\}} x_i r_{i3} + x_3 \tilde{\zeta}_3 + x_4 \tilde{\eta}_3 + \delta_3^+(\tilde{\zeta}, \tilde{\eta}) - \delta_3^-(\tilde{\zeta}, \tilde{\eta}) = 40 \quad (75)$$

$$x_i \geq 0 \quad \forall i \in I \quad (76)$$

$$\delta_{s^n}^-, \delta_{s^n}^+ \geq 0 \quad \forall s^n \in S^n \quad (77)$$

Discrete equivalents of the random variables must be generated. As there are two random variables in the constraints, two sets of random variables $R_1 = \{1, 2, 3\}$ and $R_2 = \{1, 2, 3\}$ are generated. Three bins have been chosen to reduce the number of constraints in the formulation, noting that increasing the number of bins will yield a more accurate solution.

$$\max z = \sum_{i \in I} \sum_{s^t \in S^t} x_i r_{is^t} - \sum_{r_1=1}^3 \sum_{r_2=1}^3 p_1^{r_1} p_2^{r_2} (2\delta_1^{-,r_1,r_2}) \quad (78)$$

Subject to

Prob	Impact	Variable	Bin
0.0777	4.9644	$\zeta_1^{r_1}$	1
0.8061	8.6638	$\zeta_1^{r_1}$	2
0.1162	12.3631	$\zeta_1^{r_1}$	3
0.1487	5.2685	$\eta_1^{r_1}$	1
0.7802	8.1425	$\eta_1^{r_1}$	2
0.0711	11.0164	$\eta_1^{r_1}$	3
0.1319	5.9366	$\zeta_2^{r_1}$	1
0.7789	13.3082	$\zeta_2^{r_1}$	2
0.0892	20.6799	$\zeta_2^{r_1}$	3
0.0781	3.4631	$\eta_2^{r_1}$	1
0.7801	3.9389	$\eta_2^{r_1}$	2
0.1418	4.4147	$\eta_2^{r_1}$	3
0.1140	7.0992	$\zeta_3^{r_1}$	1
0.7754	11.9197	$\zeta_3^{r_1}$	2
0.1106	16.7401	$\zeta_3^{r_1}$	3
0.1065	7.9167	$\eta_3^{r_1}$	1
0.7765	12.1088	$\eta_3^{r_1}$	2
0.1170	16.3008	$\eta_3^{r_1}$	3

Table 3

$$\sum_{i=\{1,2,5,6\}} x_i r_{i1} + x_3 \zeta_1^{r_1} + x_4 \eta_1^{r_2} + \delta_1^{+,r_1,r_2} - \delta_1^{-,r_1,r_2} = 40 \quad \forall r_1 \in R_1, \forall r_2 \in R_2 \quad (79)$$

$$\sum_{i=\{1,2,5,6\}} x_i r_{i2} + x_3 \zeta_2^{r_1} + x_4 \eta_2^{r_2} + \delta_2^{+,r_1,r_2} - \delta_2^{-,r_1,r_2} = 40 \quad \forall r_1 \in R_1, \forall r_2 \in R_2 \quad (80)$$

$$\sum_{i=\{1,2,5,6\}} x_i r_{i3} + x_3 \zeta_3^{r_1} + x_4 \eta_3^{r_2} + \delta_3^{+,r_1,r_2} - \delta_3^{-,r_1,r_2} = 40 \quad \forall r_1 \in R_1, \forall r_2 \in R_2 \quad (81)$$

$$x_i \geq 0 \quad \forall i \in I \quad (82)$$

$$\delta_{s^n}^-, \delta_{s^n}^+ \geq 0 \quad \forall s^n \in S^n \quad (83)$$

Code snippets

```
normDiscrete <- function(mean, sd, bins){
  smpl <- rnorm(n=10000, mean=mean, sd=sd)
  hist <- hist(x = smpl, breaks = seq(min(smpl),max(smpl),(max(smpl)-min(smpl))/bins))
  Probability <- hist$counts/10000
  Impact <- hist$mid
  solution <- data.frame(Probability,Impact)
  return(solution)
}
```

```
radiation.data <- read.csv(file = "RadiationData.csv", sep = ";")
```

```
data.prob <- data.frame()
```

```

# Calculating the probability and impact for zeta
for (i in 1:6){
  temp <- normDiscrete(mean = as.numeric((summarise_all(radiation.data,"mean"))[i]),
                      sd = as.numeric((summarise_all(radiation.data,"sd"))[i]),
                      bins = 3)
  data.prob <- rbind(data.prob,
                    cbind(temp,Var=names((summarise_all(radiation.data,"mean"))[i])))
}

```

Question 2 f

Model formulation

$$x_i \triangleq \begin{cases} 1 & \text{if beam } i \text{ is used, where } i \in I \text{ and } I = \{1, \dots, 6\} \\ 0 & \text{otherwise} \end{cases}$$

$$r_{is^n} \triangleq \text{radiation from beam } i \text{ hitting normal region } s^n, \text{ where } s^n \in S^n \text{ and } S^n = \{1, 2, 3\}$$

$$r_{is^t} \triangleq \text{radiation from beam } i \text{ hitting tumour region } s^t, \text{ where } s^t \in S^t \text{ and } S^t = \{1, 2, 3\}$$

$$\delta_{s^n}^- \triangleq \text{excess levels of radiation hitting normal region } s^n, \text{ where } s^n \in S^n.$$

$$\delta_{s^n}^+ \triangleq \text{reduced levels of radiation hitting normal region } s^n, \text{ where } s^n \in S^n.$$

The fifth and fourth beams are denoted as random variables such that $\tilde{r}_{3s^n} = \{\tilde{\zeta}_1, \tilde{\zeta}_2, \tilde{\zeta}_3\}$ and $x_4 \tilde{r}_{4s^n} = \{\tilde{\eta}_1, \tilde{\eta}_2, \tilde{\eta}_3\}$. In addition to the variables included in the recourse approach, a new variable is introduced to determine the reliability of the model.

$$r_n \triangleq \begin{cases} 1 & \text{if either } z^{\text{actual}} \leq z^*, \text{ or } \delta_{s^n}^- + \delta_{s^n}^+ = 0 \\ 0 & \text{otherwise} \end{cases}$$

The reliability may then be calculated as

$$\frac{1}{N} \sum_{n \in \{1, \dots, N\}} r_n \tag{84}$$

A quick comparison between the expected value approach and the analytical approximation approach reveals that the expected value approach obtained a smaller difference between the maximum radiation levels for normal tissue, leading it to be less conservative. By deduction, the scenario approach is the most constraining and thus caters for greater spread of possibilities.