Exercise #4: Determining the optimal height of flood protection structures: the van Dantzig (1956) example

Patrick Applegate, patrick.applegate@psu.edu

Learning Goals

After completing this exercise, you should be able to

- describe the van Dantzig (1956) approach to estimating the optimal height of flood protection structures
- explain how discounting works and describe the effects of discounting rates on the results of cost-benefit analyses
- perform a simple Monte Carlo analysis based on van Dantzig (1956)

Introduction

Suppose that we are aware of a potential future danger. We don't know the magnitude of this danger, but we do know the probability associated with different levels of danger. We also know the costs of putting protection in place. These protective measures are imperfect, and we don't want to spend more money on protection than is necessary. How do we decide how much protection to institute?

van Dantzig (1956) considered this problem in the context of avoiding damaging floods from storm surges in the Netherlands. The Netherlands is a densely-populated and low-lying country in Europe. Storm surges occur when low-pressure regions associated with large storms cause local sea level to rise temporarily. These storm surges can be several meters high. A particularly large and damaging flood occurred in the Netherlands in 1953.

After the flood of 1953, a team of engineers and mathematicians set to work to design a set of dikes (levees) to protect the Netherlands from similar future events. It was apparent that the dikes could not be made tall enough to close out all potential future floods. The floods that had already happened suggested that even bigger ones could occur. As dikes are built higher, they also become wider and take up valuable real estate. Thus, the cost of building very tall dikes becomes unreasonably large.

van Dantzig (1956) suggested that the height of the dikes should be increased until the sum of their construction cost and the damages expected from future floods reaches a minimum. This criterion is economically optimal *if*

- the construction costs of the dikes as a function of their height,
- the probabilities of different flood magnitudes, and
- the damages associated with flooding

are well-known.

Dike construction costs

van Dantzig (1956) argued that the cost of building a dike I is a linear function of height H,

$I(H) = I_0 + kH$,

for small values of H. Strictly speaking, van Dantzig (1956) analyzed the case in which dikes with an initial height H_0 are increased to a final height H, with the height increase $X = H - H_0$. Here, we assume that there are no existing protective structures, so that $H_0 = 0$ and $H = X$.

Expected damages

The question is then, how much damage should we expect from floods if we build the dikes to a particular height? According to van Dantzig (1956), we should

1. calculate the expected damage in a single year and
2. integrate these damages over all future times,
3. while assigning a reduced importance to floods that happen far in the future.

The **expected damage in a single year** is the product of the losses due to a single flood, multiplied by the chances of a flood in that year. van Dantzig (1956) assumed that, in the case of a flood, the value V of all goods (buildings, livestock, and other property) within the area protected by a dike would be lost. From existing sea level data, van Dantzig (1956) knew that the probability that flood waters will overtop a dike of a given height in a particular year follows an exponential distribution in the northern Netherlands,

$P(H) = p_0 e^{-\alpha H}$

Recall from Exercise #3 that an exponential distribution assigns the largest probabilities to small values, but has a long "tail" that can extend to large positive values. Multiplying the damage from one flood by the probability of flooding in one year gives

$VP(H) = Vp_0 e^{-\alpha H}$

But, if we invest money now and it grows at an interest rate δ, we will have more money in the distant future, and we'll be better able to pay for flood damages that occur then. This *discounting* **assigns a reduced weight to floods that happen far in the future**, and is a common feature of cost-benefit analyses. So, we define a *discounting factor* F_d that assigns a weight to future potential floods depending on how many years t have gone by when they happen,

$F_d = (1 + \delta)^{-t}$

Putting the last two equations together, van Dantzig (1956) arrived at an expression for the total future losses L as a function of dike height H. The summation (Σ) in this expression **integrates flood damages over all future times**.

$L(H) = VP(H) \sum_{t=0}^{\infty} F_d$

Substituting for $P(H)$ and F_d and simplifying, van Dantzig (1956) obtained

$L(H) \approx (Vp_0 e^{-\alpha H})/\delta$

How could van Dantzig (1956) perform this simplification? Performing the summation $\sum_{t=0}^{\infty}(1+\delta)^{-t}$, we see that it tends to $(1 + \delta)/\delta$. For small values of δ, this factor is close to $1/\delta$. The code block below generates a figure (Fig. 1) that demonstrates this point.

```
# Demonstration that the cumulative sum of the discount factor tends to
# (1+ delta)/ delta as t becomes large, and that this value is close to
# 1/ delta for small values of delta.
delta <- 0.04              # discounting rate
t <- seq(0, 200, by = 1)   # vector of time values (years)
F.d <- (1+ delta)^ -t      # discounting factor as a function of time
plot(t, cumsum(F.d), type = 'l', bty = 'n', lwd = 2, xlab = 'Time (yr)',
     ylab = 'Running total, discount factor', col = 'blue')
abline(h = (1+ delta)/ delta, lty = 2, lwd = 2, col = 'red')
abline(h = 1/ delta, lty = 2, lwd = 2, col = 'gray')
```

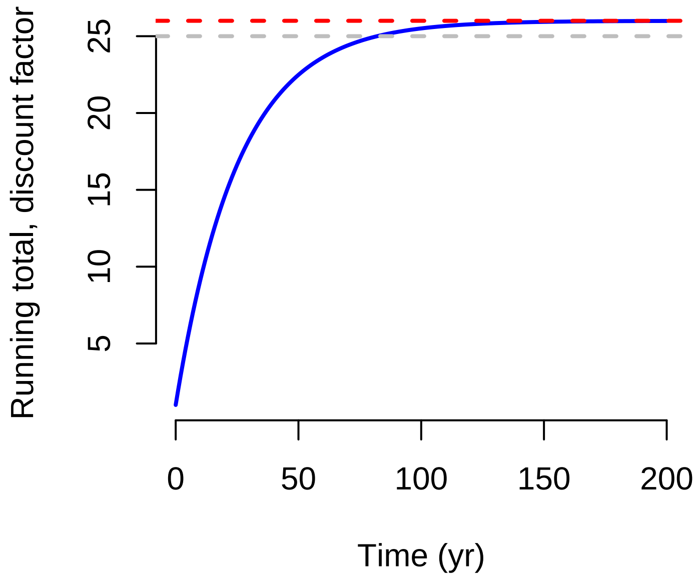


Figure 1: Cumulative sum of the discount factor as a function of time, for a discounting rate $\delta = 0.04$. This cumulative sum tends toward $(1+\delta)/\delta$ (red, dashed line) as t becomes large. This value is close to $1/\delta$ for small values of δ (gray, dashed line).

Finding the optimal dike height

Recall that van Dantzig (1956) suggested that the height of the dikes should be increased until the sum of their construction cost and the damages expected from future floods reaches a minimum. We now have an equation that describes the cost of building dikes as a function of their height, $I(H) = I_0 + kH$, and another equation that describes the expected future damages due to floods, also as a function of dike height, $L(H) \approx (Vp_0e^{-\alpha H})/\delta$.

The following code block plots $I(H)$, $L(H)$, and their sum $I(H) + L(H)$, and finds the minimum on the total curve. This point corresponds to the optimal dike height.

```
# Constants from van Dantzig (1956)
p_0 = 0.0038     # unitless; probability of flooding in a given year if the dikes
                 # aren't built
alpha = 2.6      # unitless; constant associated with flooding probabilities
V = 10^ 10       # guilders; value of goods threatened by flooding
delta = 0.04     # unitless; discount rate
I_0 = 0          # guilders; initial cost of building dikes
k = 42* 10^ 6    # guilders/m; cost of raising the dikes by 1 m

# Make a vector of possible dike height increases in meters.
H = seq(0, 3, by = 0.001)

# Calculate the expected losses due to floods L, the cost of increasing the
# dikes I, and the total of L and I, all as a function of H.
L = p_0* exp(-alpha* H)* V/ (delta)
I = I_0+ k* H
Total = L+ I

# Make a plot.  The lowest point on the Total curve
# is the best value for H.
```

```r
plot(H, L, type = 'l', col = 'red', xlab = 'Dike height increase H (m)',
     ylab = 'Cost (guilders)', bty = 'n')
lines(H, I, col = 'blue')
lines(H, Total, col = 'black', lwd = 2)

# Add a point indicating the minimum.
min_Total = Total[rank(Total) == 1]
min_H = H[rank(Total) == 1]
points(min_H, min_Total, pch = 16)
```

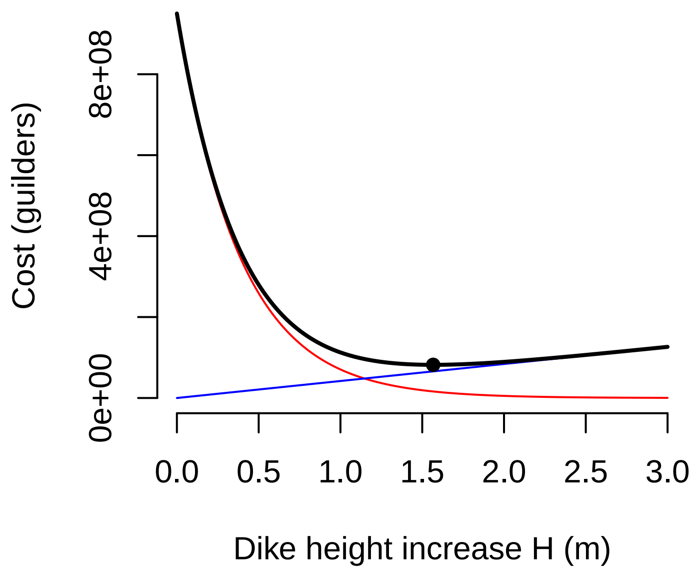


Figure 2: $I(H)$ (blue curve), $L(H)$ (red curve), and their sum (black curve). The minimum point on the black curve indicates the optimal dike height.

Monte Carlo simulation

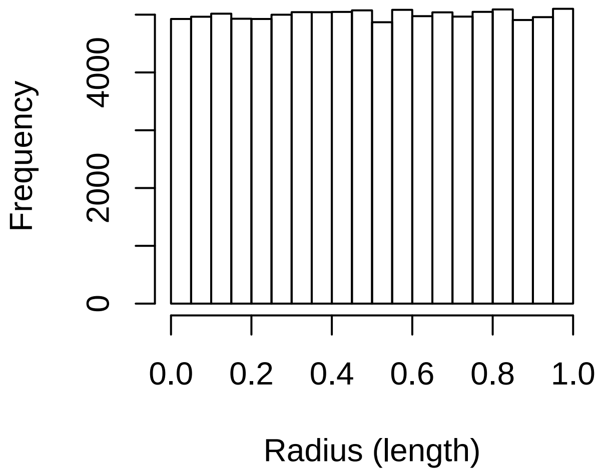
The method proposed by van Dantzig (1956) for estimating optimal dike heights includes a number of parameters whose values have to be estimated (see the list in the code block above). Suppose that the probability distribution of floods of different heights is well-known and the initial cost of building higher dikes is 0 (that is, `p_0`, `alpha`, and `I_0` are fixed). In that case, we still have three uncertain parameters, the value of goods protected by flooding `V`, the discount rate `delta`, and the cost of raising the dikes by 1 m `k`. How can we assess the uncertainty in the optimal dike height, given that these parameters aren't known perfectly?

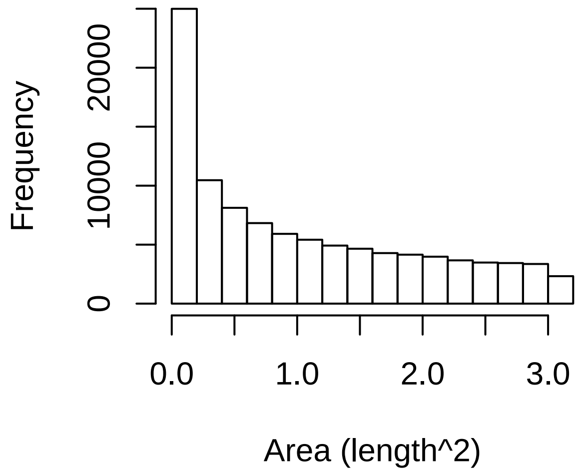
Monte Carlo simulation (e.g. Bevington and Robinson, 2002, their ch. 5) provides a method for estimating the uncertainty in a calculated output, given probability distributions of the inputs. In Monte Carlo simulation,

1. groups of input parameter values are generated randomly using the techniques described in Exercise #3,
2. each of the groups of input parameter values is fed into the model and the outputs are recorded, and
3. the distributions of the output are plotted.

As a simple example, suppose we want to estimate the area of a circle, but we only know that its radius is somewhere between 0 and 1 units. We assume that the distribution of radius values is uniform, make a vector of possible values, feed them into the equation for a circle, and histogram the results. As shown in Figure 3, the distribution of outputs looks very different from the distribution of inputs.

```r
# Creates randomly-sampled values of a circle's radius, calculates the area of
# the circle from each of the radius values, and makes histograms of the
# radius and area values.
n.trials <- 10^5    # number of Monte Carlo trials to attempt
radius <- runif(n.trials, min = 0, max = 1)
area <- pi* radius^2
hist(radius, main = '', xlab = 'Radius (length)')
hist(area, main = '', xlab = 'Area (length^2)')
```





In the code block above, note that both `radius` and `area` are vectors that each contain `n.trials` values.

How could we apply Monte Carlo simulation to the van Dantzig (1956) analysis? In addition to the equations above, van Dantzig (1956) presents a separate equation that gives the optimal dike height directly,

$H_{best} = \alpha^{-1} \ln[(V p_0 \alpha)/(\delta k)]$

We can imagine generating vectors of the uncertain parameters, feeding them into the equation above, and making a histogram of the optimal dike height values for all the parameter groups. That histogram would give us some idea of how sure we can be about the optimal dike height.

Tutorial

Open the R script `lab4_sample.R` and inspect the contents. This script performs a simple Monte Carlo analysis with van Dantzig (1956)'s equation for the optimal dike height, varying the parameter δ between 0.95 and 1.05 of its base value.

The first part of the code is straightforward; it provides an explanation of what the script does, clears the workspace, and sets the values for different parameters in the equation above.

```r
# lab4_sample.R
# Patrick Applegate, patrick.applegate@psu.edu
#
# Performs a simple Monte Carlo analysis with the optimal dike height
# equation from van Dantzig (1956).

# Clear any existing variables and plots.
rm(list = ls())
graphics.off()
```

```r
# Constants from van Dantzig (1956)
p_0 = 0.0038      # unitless; probability of flooding in 1 yr if the dikes
                  # aren't built
alpha = 2.6       # unitless; constant associated with flooding probabilities
V = 10^ 10        # guilders; value of goods threatened by flooding
delta = 0.04      # unitless; discount rate
I_0 = 0           # guilders; initial cost of building dikes
k = 42* 10^ 6     # guilders/m; cost of raising the dikes by 1 m
```

The next group of commands determines how many Monte Carlo calculations to perform (`n.trials`), sets the range of each parameter to search over (`range`), and affects how wide a range of values will be reported by the script when it's run (`probs`). `probs <- c(0.025, 0.5, 0.975)` tells the script to report the 95% range of the results, plus the median.

```r
# Set some other values.
n.trials <- 10^5  # number of Monte Carlo trials to do
range <- 0.1      # fractional range of each parameter to test
probs <- c(0.025, 0.5, 0.975)
                  # which quantiles to report
```

The random sampling is handled in the next block of code. Note the use of the `set.seed()` command to ensure that the script will give reproducible results.

```r
# Set the seed for random sampling.
set.seed(1)

# Perform the random sampling.
facs <- c((1- 0.5* range), (1+ 0.5* range))
delta.vals <- runif(n.trials, min = facs[1]* delta,
               max = facs[2]* delta)
```

Finally, the code calculates the optimal dike height for each value of `delta.vals` and makes a histogram of the `best.heights`, with a vertical red line to indicate the height obtained using the best estimate of each uncertain parameter. The quantiles of the values in `best.heights` are also written to the screen.

```r
# Calculate the optimal dike heights.
best.heights <- alpha^-1* log((V* p_0* alpha)/ (delta.vals* k))

# Make a histogram and print the quantiles to the screen.
hist(best.heights, main = '', xlab = 'Optimal dike heights (m)')
abline(v = alpha^-1* log((V* p_0* alpha)/ (delta* k)), lwd = 2, col = 'red')
print(round(quantile(best.heights, probs = probs), 3))
```

Exercise

Part 1. Make a plot of the discounting factor F_d for `delta = seq(0, 0.1, by = 0.2)` over the time interval 0-200 yr. (Make sure you are plotting the discounting factor, not its cumulative sum as shown in Fig. 1.)
Part 2. Execute `lab4_sample.R` and take note of the quantiles that the script produces. Save a copy of the histogram produced by the script.
Part 3. Make a copy of `lab4_sample.R` by saving it with a different file name. Modify this copied file so that it incorporates randomly-selected `V` and `k` values into the Monte Carlo simulation. You'll need to

create vectors `V.vals` and `k.vals` and populate them with random values, using code similar to that for `delta.vals`, above. You'll also need to change the line in which `best.heights` is calculated, to incorporate these values into calculation of the optimal dike heights. Execute this new script, take note of the quantiles it produces, and save a copy of the histogram it produces.

Questions

1. For each of the discount factors you investigated in Part 1, how much weight do losses at the end of 100 years have relative to losses now?
2. Compare the distribution of `best.heights` from Part 2 to the distribution of values in `delta.vals`. Do these distributions look like one another?
3. Now compare the histograms and quantiles from Parts 2 and 3 to one another. How does the distribution of `best.heights` change as more free parameters are added to the calculation?

References

Bevington, P. R., and Robinson, D. K., 2002. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, 320 p.

van Dantzig, D., 1956. Economic decision problems for flood prevention. Econometrica 24, 276-287.