Thinking Probabilistically II

Vik Gopal

2019-01-21

It's All In The Game[1]

[1] Nat King Cole, 20 Golden Greats (1978)

Decisions for the Decade

recap

- Let us consider the first decade to begin with.
- Every year, **exactly one** of three things happens:
  - Flood,
  - Drought, or
  - No extreme event
- The probability of flood was equal to the probability of drought, which was $1/6$.
- We had 10 beans to decide what to do with:
  - Protect against flood,
  - Protect against drought, or
  - Keep as an investment.
- Every extreme event that happened used up a protection bean.
- If we have no more protection beans when an extreme event happens, we lose all our investment beans too.

Extreme Events Model

random variables

We introduce the following random quantities:

$$
\begin{aligned}
N_f &:= \text{the number of floods in a decade} \\
N_d &:= \text{the number of droughts in a decade} \\
N_o &:= \text{the number of ordinary years in a decade}
\end{aligned}
$$

- These are the counts that we need in order to compute our final investment return, *at the end of each decade*.
- At the beginning of the decade, we will not know their precise values.

Extreme Events Model

- We need a representation of the probabilistic behaviour of the game in slide 3.
- The abstract description is this:
  - We have three labelled bins, each having a different probability associated with it.
  - We have 10 numbered balls to put into them.
  - We pick one bin according to the following probabilities, and place ball 1 into it.
    - The first and second bins will be picked with probability $1/6$, and the third bins will be picked with probability $2/3$.
  - Repeat for balls 2 to 10.
- We can represent our outcome as a 3-tuple. For instance, if we observed 3 floods, and 2 droughts, then we can write

$$(N_f, N_d, N_0) = (3, 2, 5)$$

Model Suitability

pause

All models are wrong, but some are useful.

– George Box, 1978

Take a few moments to write down 2 things that you feel are unrealistic about this model.

Decision Space

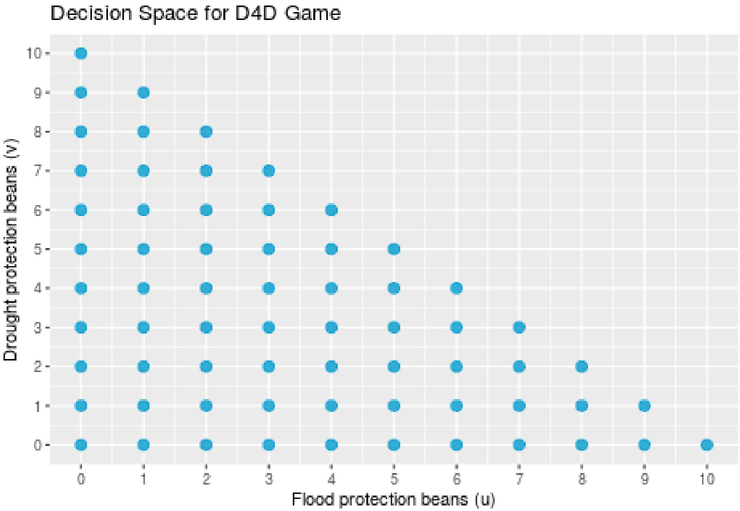
- Our ultimate goal is to make a decision.
- The space of possible decisions varies from problem to problem.
- In our case, we have 10 beans, and we need to decide how many to put up for flood and drought protection, and how many to put up for investment.
- Suppose we let u and v represent the number of beans for flood and drought protection respectively. We do not have an unrestrained choice of u and v. These are some of the constraints:
  - u and v can only be between 0 and 10.
  - The sum of u and v must be 10 or less.
- u and v are not random. They are decided by us at the beginning of the decade.
- We could represent our choice of u and v with a tuple. For instance,

$$(u, v) = (5, 3)$$

  would mean we spent 5 beans on flood protection, 3 beans on drought protection and 2 beans on investment.

Decision Space

visualisation



Investment Return

aka negative loss function

- We make our decision at the beginning of the decade, but the consequence (investment return) is
  - random, and
  - only known at the end of the decade.
- How can we represent the return in terms of the decision and the random variables introduced earlier?
- Let's start with the basic components:
  (a) $10 - u - v$ beans are reserved for investment. This is the maximum amount we can get back at the end of the decade.
  (b) If we **do not have** enough protection, a crises results. This translates to
    - ★ If $N_f > u$, we have a flood crisis; all $10 - u - v$ beans are lost.
    - ★ Similarly, if $N_d > v$, we have a drought crisis.

  (c) If we **do indeed have** enough protection for the extreme events that transpired, then we get to keep the $10 - u - v$ beans.

Expected Investment Return

- To compute the expected return, we follow the approach in the earlier session:
- Suppose we set $u = 4$ and $v = 3$.

N_f	N_d	N_o	Chance	Return	Product
4	2	4	p_1	3	$3p_1$
3	1	6	p_2	3	$3p_2$
6	2	2	p_3	0	0
⋮	⋮	⋮	⋮	⋮	⋮

The expected return is the sum of the right-hand most column.

- Notice that the expected return depends on what u and v values we set. The expected return is a **function of our decision** alone.
- If we can find a (u, v) pairing that maximises this, we can have some justification in our actions.

Loss Function Suitability

pause

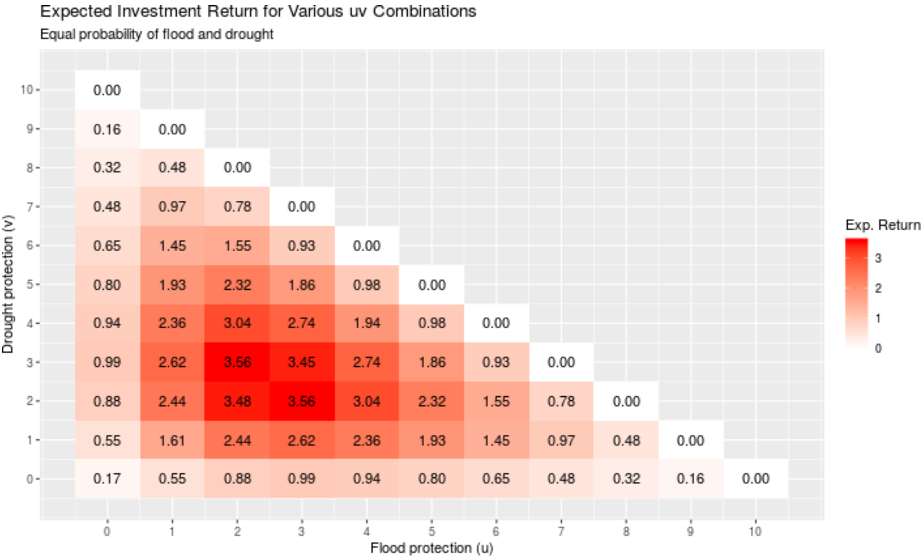
> *Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.*
>
> *– John W. Tukey, 1962*

Take a few moments to write down 2 things that you feel are unrealistic about this loss function.

Always Trying to Work It Out[2]

[2] Low, Double Negative (2018)

Decade 1 Game



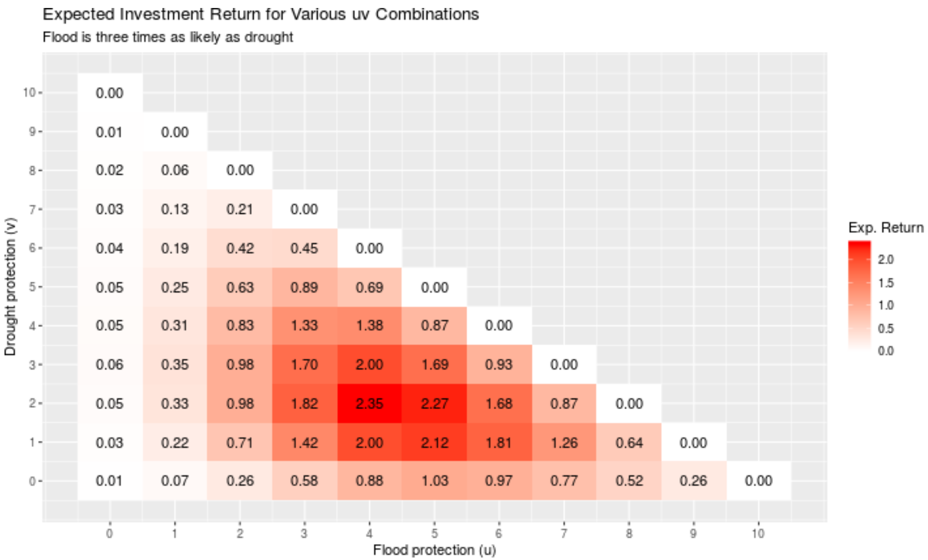
Decade 2 Game Specs

- In the second decade of the game, we used a different set of assumptions.
- We used the knowledge that the probability of flood is now three times the probability of drought.
- The new probabilities were

$$(\underbrace{3/8}_{\text{flood}}, \underbrace{1/8}_{\text{drought}}, 4/8)$$

(ref. slide 5).

Decade 2 Game



Review

- Remember how you wrote down some improvements to the model?
- Let us make a list of them now.

Some Days Are Better Than Others[3]

[3] U2, Zooropa (1993)

New Elements in Return Function

- First of all, let us suppose we have 10 million dollars to invest in rather than 10 beans.
- Then, suppose that, instead of keeping our money under the bed, we put it in a hedge fund, that returns us 5% a year.
- We pay $1 million as insurance against each extreme event. However, if we are unprotected, we will have to fork out $2 million for each unprotected flood and $3 million for each unprotected drought.
- The new return can be expressed (in English) as follows:

$$\text{investment return} + \text{flood return} + \text{drought return}$$

- The expected value can be computed one term at a time, and then summed up.

Flood/Drought Return

- Let us consider the portion of return related to floods.
- If the observed $N_f \leq u$, we do not have to pay any damages.
- On the other hand, if $N_f > u$, we have to pay $2(N_f - u)$ million dollars.
- A succint mathematical representation of this is

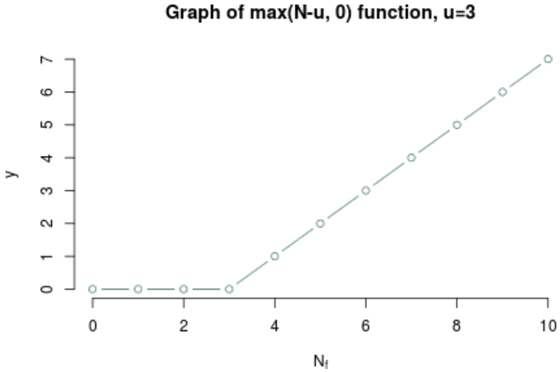
$$-2\max(N_f - u, 0)$$

- Similarly, the return from drought can be represented as

$$-3\max(N_d - v, 0)$$

Graph of Damage Function

The max$(\cdot, \cdot)$ function is a little difficult to grasp first time around; here is a graph of it to help:



There is non-zero damage only when there were more floods than protection.

Expected Investment Return

- For the investment portion, we would make the following amount:

$$(10 - u - v)(1.05)^{10}$$

- The full return amount is thus:

$$(10 - u - v)(1.05)^{10} - 2\max(N_f - u, 0) - 3\max(N_d - v, 0)$$

- What is random and what is fixed by us here?
- Compare the new return to the old one, in the first two decades of the game (ref. slide 10).
- How do we compute the expected value now???

Computing Expected Value

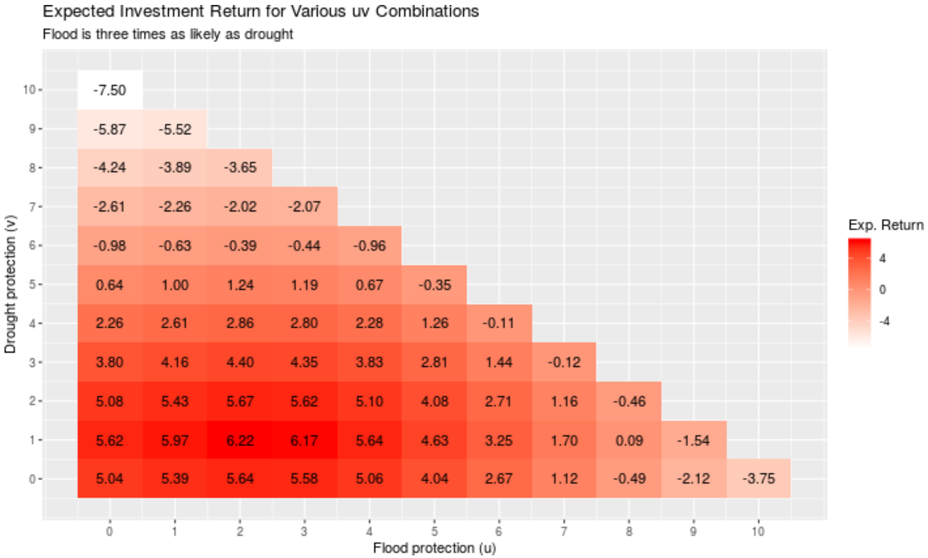
flood return

- Remember that expectation is a sum of "possibilities times probabilities", and we can compute it term by term.
- For instance, for $u = 4$, the expected flood return is computed as follows:

N_f	N_d	N_o	Chance	$-2\max(N_f - u, 0)$	Product
4	2	4	0.061	0	0.000
3	1	6	0.087	0	0.000
6	2	2	0.014	-4	-0.055
⋮	⋮	⋮	⋮	⋮	⋮

- Once again, the expected damage from flood is the sum of the column on the right.

Expected Return, New Function

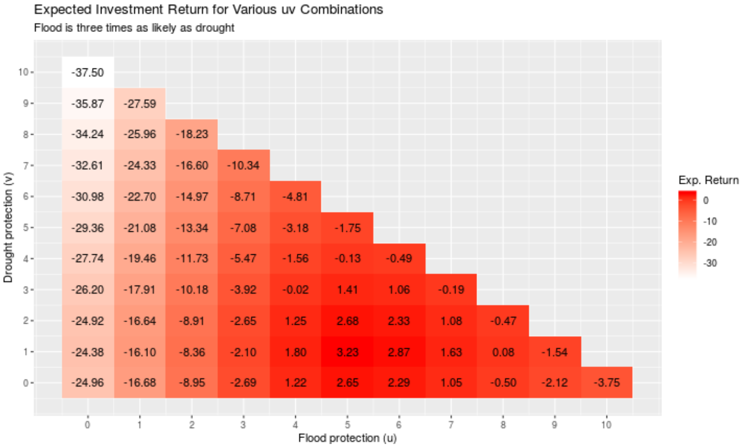


Expected Return, New Function

alternate damage specification

Suppose the damage from flood was more severe than we put down:

$$(10 - u - v)(1.05)^{10} - 10\max(N_f - u, 0) - 3\max(N_d - v, 0)$$



2 Rights Make 1 Wrong[4]

[4] Mogwai, Special Moves (2010)

Misspecifications

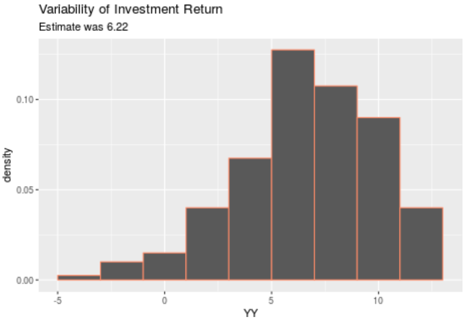
- What if the true cost function should be the one on slide 25, but we used the one on slide 22?
  - We would decided $(u = 2,\ v = 1)$ instead of $(5, 1)$.
- What if we had computed class probabilities wrongly?
- What if class probabilities were changing over time?

Protection Against Model Misspecification

- Use historical data to estimate parameters.
- Track and update model regularly.
- Run simulations to assess the uncertainty in our estimation of investment return.

Variability of Returns

- We have made a decision using the **expected return**. In truth, the return is a random variable.
- We hope to do well in the long run, but we could be unlucky on any particular decade.



- The above histogram is for the cost function on slide 22.

Evan Finds The Third Room[5]

[5] Khruangbin, Con Todo El Mundo (2018)

Decade 3 Game

- In the third decade, we did not know what the probabilities of the extreme events was.
- How can we fit this into our framework?
- We still need some way of representing the situation in mathematical terms.
- Perhaps we say

$$\begin{aligned} \text{probability of drought} &\in (0.1, 0.2) \\ \text{probability of flood} &= 2 \times \text{probability of drought} \end{aligned}$$

- The formal decision theoretic approach requires a Bayesian formulation of the problem.
- Essentially, we are still after the expected loss, but the probabilities are now random variables.

Ready, Steady, Go[6]

[6] Paul Oakenfold, Bunkka (2002)

Model-Building Tips

... premature optimisation is the root of all evil.

– Donald Knuth.

- No model is perfect. Don't try to get the perfect model the first time. Start with simple approximations.
- Model building is an iterative process.
  (i) Start with historical data to estimate parameters and verify the model.
  (ii) Track the performance. Assess goodness-of-fit regularly.
  (iii) Study sensitivity of solution, variability of estimates. We want robust solutions (decisions) as far as possible.
  (iv) Analyse the residuals, go back and alter the model if necessary.