



A Resilient Future: Science and Technology for Disaster Risk Reduction

Professor Anthony Davison
Chair of Statistics, STAT at EPFL



Search MOOC



Video



EPFL



- Introduction
- What do we mean by risk?
 - Random experiment
 - Expected loss
- How do we estimate risk?
 - Fitted probability model
 - Loss function
- Variation and uncertainty
- Main points

A Resilient Future: Science and Technology for Disaster Risk Reduction

Hello. My name is Anthony Davison, I'm Professor of Statistics here at EPFL, the École polytechnique fédérale de Lausanne, and I have a particular interest in trying to estimate the probabilities of rare events. The purpose of this brief video, is to give an introduction to the idea of risk, to the idea of a probability model, which involves the idea of a random experiment, and then to discuss briefly how we use probability models, and some idea of losses, to estimate potential risks.

Notes

Summary



0m 05s

Introduction to the definition of risk



A Resilient Future: Science and Technology for Disaster Risk Reduction

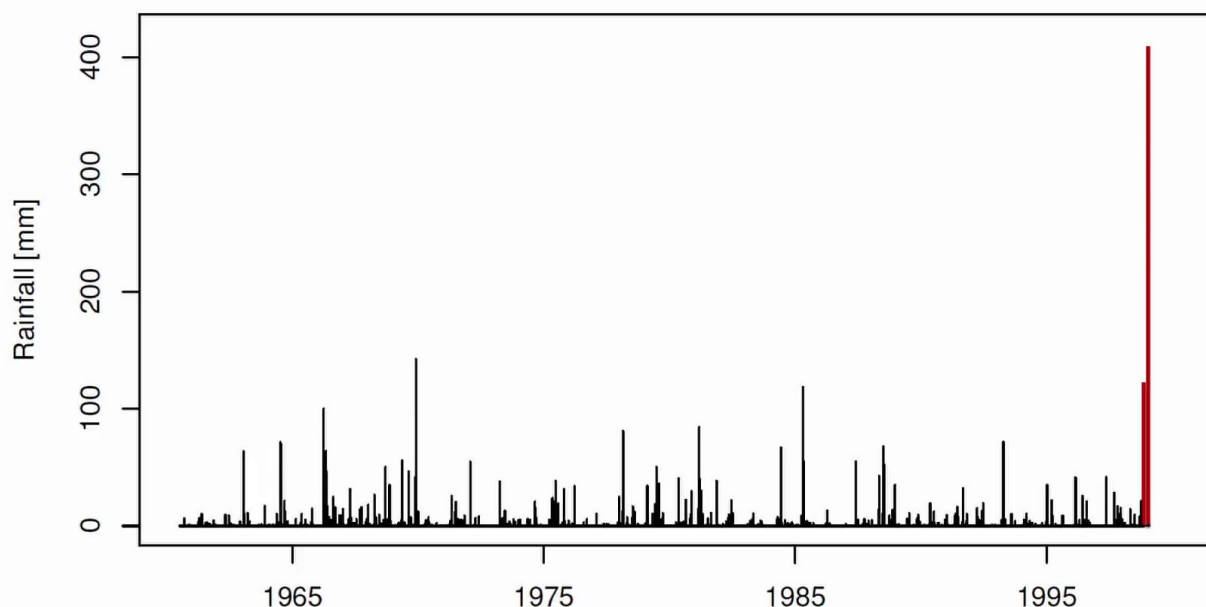
Let's begin by looking at a tragic case study. The state of Vargas is a state on the northern coast of Venezuela, which you can see here in the map. And it's got a very mountainous terrain, as you can see in the picture here on the right. When it rains heavily here, it's very rapid runoff, and that can lead to landslides.

Notes

Summary



Introduction to the definition of risk



(5)

A Resilient Future: Science and Technology for Disaster Risk Reduction

This graph shows data on daily rainfall amounts, from the beginning of 1961 up to almost the end of 1999, at Maiquetia, the airport near Caracas. You can see there's a lot of variation, as there always is in daily rainfall data, with the largest value over that period, being about 150 millimeters in a given day. The month of December, 1999, had some quite heavy rainfall. And then in the middle of the month, 14th, 15th and 16th, massive amounts of rainfall fell, about 90 centimeters over the three days. And that led to enormous flooding, runoff and landslides and mudflows.

Notes

Summary



1m 00s

What do we mean by risk?



- The expected loss due to an undesirable outcome in a random experiment
- Key words:
 - random experiment
 - outcome
 - expected loss

A Resilient Future: Science and Technology for Disaster Risk Reduction

What you see in the picture here, is the result of a landslide on what was a bustling town, and you can see it's been devastated. No one knows how many people lost their lives. Estimates range from 10,000 to 30,000, but no one really knows, because there was no population census of any sort. So what do we mean by risk? The definition of risk is the expected loss due to an undesirable outcome in a random experiment. And there, there are three key words: random experiment, outcome and expected loss.

Notes

Summary



1m 52s

What do we mean by risk? Random experiment



The total probability is 1, because one of the outcomes must occur when the experiment is performed

Outcome	Probability
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

So let's think what we might mean by a simple random experiment, where I take this die and throw it once. There are six possible outcomes, one up to six. And if we suppose that the die was fair, then their respective probabilities would be 1/6 each, and then we would have this table here, of outcomes with the corresponding probabilities. The total probability is 1, because exactly one of those outcomes must occur when the probability's performed.

Notes

Summary



What do we mean by risk? Random experiment



The event “the outcome is odd” corresponds to observing 1, 3 or 5, and has probability

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

since the individual outcomes cannot occur together

A Resilient Future: Science and Technology for Disaster Risk Reduction

We can then use that table of probabilities, to work out the probabilities of more complicated events, such as the event that the outcome is odd, which corresponds to observing a one, a three or a five, and because those different outcomes are disjoint, none of them can occur together, the total probability is the sum of the individual probabilities for the three outcomes, namely $1/6$ plus $1/6$ plus $1/6$, and that gives us $1/2$.

Notes

Summary



2m 59s

What do we mean by risk? Expected loss



In a moment of madness I agree to pay you 6 dollars if the die shows a 6

A Resilient Future: Science and Technology for Disaster Risk Reduction

Now, let's imagine that in a moment of madness, I agree to pay you 6 dollars if the die gives a 6, and nothing otherwise.

Notes

Summary



3m 29s

What do we mean by risk? Expected loss



My loss will be 6 dollars, if the undesirable (to me!) outcome 6 occurs, and zero otherwise:

Outcome	Probability	Loss
1	1/6	0
2	1/6	0
3	1/6	0
4	1/6	0
5	1/6	0
6	1/6	6

Then we get this table here, which gives us the outcomes, the probabilities and the potential losses.

Notes

Summary



3m 38s

What do we mean by risk? Expected loss

The outcome 6 occurs with probability $1/6$, so on average it happens only every 6 throws. Thus, the expected loss due to the undesirable outcome is:

$$(\text{probability of a 6}) \times (\text{loss when 6 observed}) = \frac{1}{6} \times 6 = 1 \text{ dollar}$$

A Resilient Future: Science and Technology for Disaster Risk Reduction

The outcome 6 occurs with probability $1/6$, so on average it occurs *only* once every 6 throws, and so the expected loss due to the particular undesirable outcome 1, of getting a 6-- undesirable for me, of course, highly desirable for the person I'm playing the game with-- is the probability of a 6, times the loss when that's observed, which is $1/6$ times 6, which is 1 dollar. So that's my expected loss when we play this game.

Notes

Summary



3m 48s

What do we mean by risk? Expected loss



My potential losses are:

Outcome	Probability	Loss
1	1/6	1
2	1/6	2
3	1/6	3
4	1/6	4
5	1/6	5
6	1/6	6

We can also imagine a more expensive game for me, in which I give you the amount in dollars shown on the face of the die. So in this case, that would be 4, for example, or 6, or whatever comes up when the die is thrown. In that case, my potential losses are from 1 up to 6, with the corresponding probabilities.

Notes

Summary



4m 21s

What do we mean by risk? Expected loss

The probability is $\frac{1}{6}$ that I lose 1 dollar, $\frac{1}{6}$ that I lose 2 dollars, etc., so my expected loss when the experiment is performed is now:

$$\frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = \frac{21}{6} = 3.5 \text{ dollars}$$

A Resilient Future: Science and Technology for Disaster Risk Reduction

And we can compute the expected loss using the formula given here, I lose 1 if the outcome is 1, the probability is $\frac{1}{6}$, 2 if the outcome is 2, with probability of $\frac{1}{6}$, and so on. That gives me an expected loss of 3.5 dollars, every time we throw the die.

Notes

Summary



4m 45s

How do we estimate risk? General approach

- In a given application, we must construct a probability model:
 - We define the random experiment
 - We define the outcomes
 - We assign a probability to each outcome
- We express the loss for each outcome in a loss function:

$$\text{expected loss} = \sum_{\text{outcomes}} (\text{loss due to outcome}) \times (\text{probability of outcome})$$

A Resilient Future: Science and Technology for Disaster Risk Reduction

So that gives us an idea how, in a particular application, we can estimate risk. The die example, of course, is almost the simplest example we could think of. In general we have to do things slightly differently, but the same basic steps apply. We have to construct a probability model, we have to define the random experiment, define the outcomes and somehow assign probabilities to each outcome. We express the loss for each outcome in a loss function, and then we can compute the expected loss, which is the sum over outcomes, of the loss due to each outcome, times the probability of that outcome.

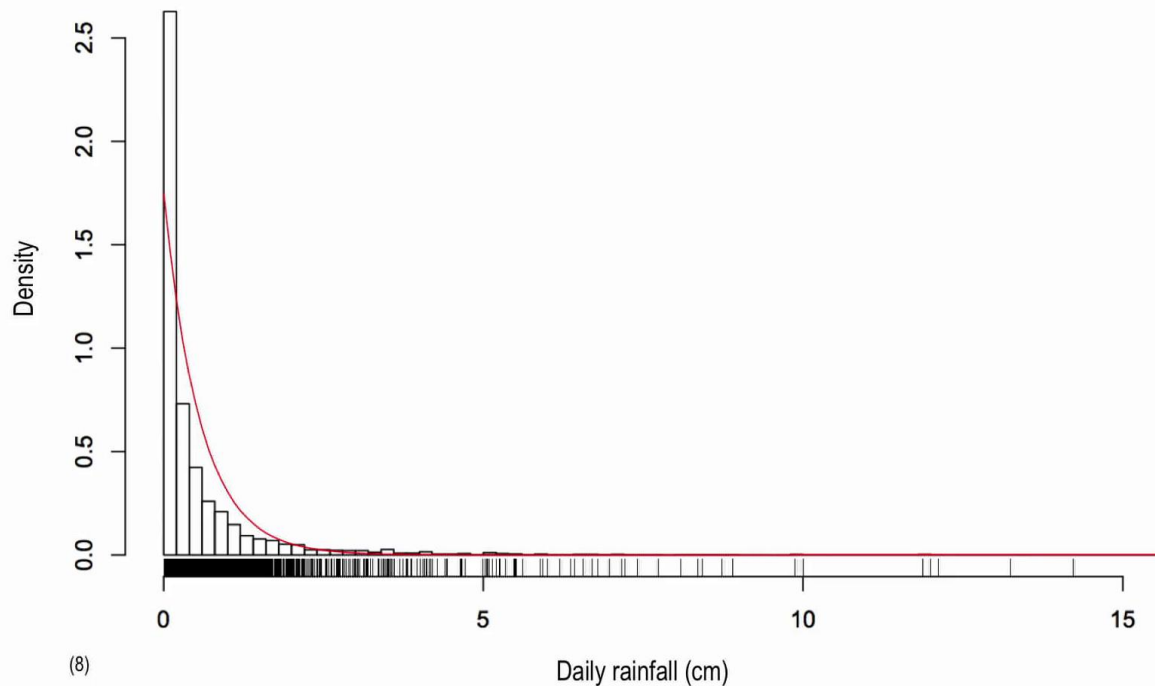
Notes

Summary



5m 10s

How do we estimate risk? Fitted probability model



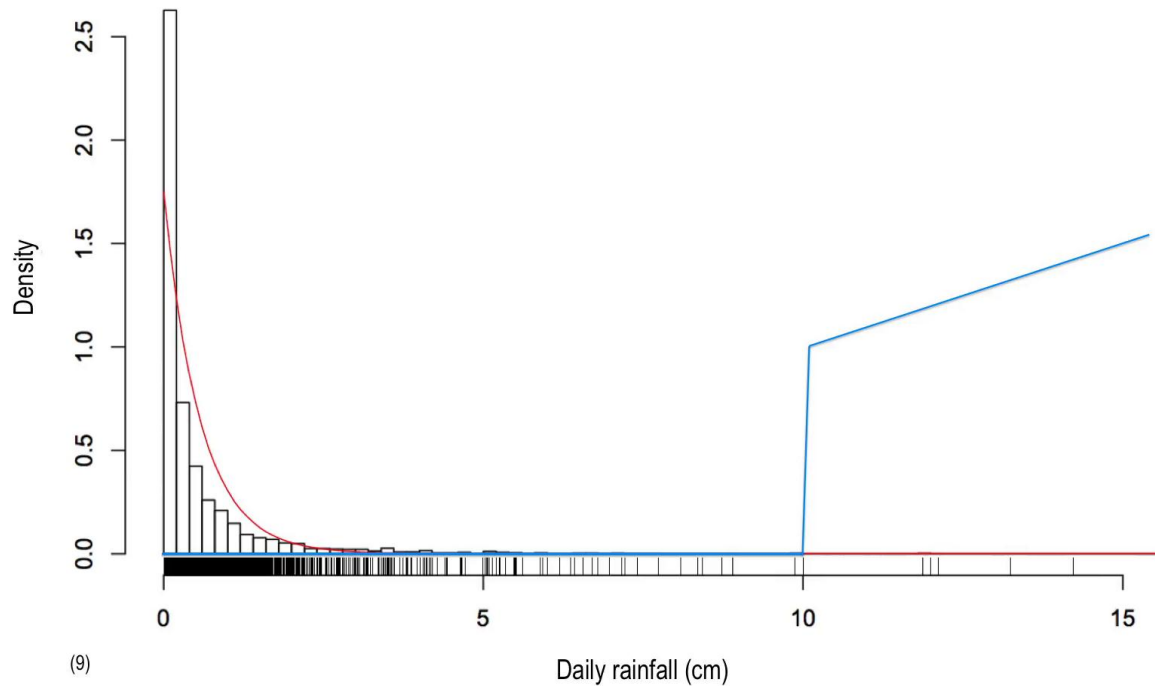
The histogram here, shows the daily rainfall amounts up to the beginning of December, 1999. You can see that the largest value is about 14.5 or so, and the question arises, how could we use those values to estimate the probabilities of the later events of that month? It's clearly not enough to use these data on themselves, because there's no value larger than 14.5 and therefore we couldn't just predict one. Also, there are values that there are holes, as you can see, we wouldn't be able to predict a value between say 11 and 12, because no previous value has been observed in that range. So what we have to do, is to replace the data by a theoretical construct that's supposed to represent them. Here's a picture showing what such a construct might look like. What I've done is to fit a probability model to the data. You can see that it's not a very good model, because it underestimates the probability of very small values, and then overestimates the probabilities of values up to about 2.5 or so, because the red line that represents the theoretical probability model, is higher than the observed values there. And of course, because the total probability is 1, that will mean that it will be underestimating the probabilities of larger events.

Notes

Summary



How do we estimate risk? FPM and loss function



But the particular line here isn't important, because what I'm trying to explain is the idea of replacing the data with a probability model. In applications, of course, we would try and choose a suitable class of probability models, that represent rare events appropriately, and that means thinking hard about the construction of such models. There is an extensive statistical and probabilistic literature on that. In addition to a fitted probability model, we also need a loss function. What this picture here shows is an example of such a loss function, where I arbitrarily suppose that the loss is 0 up to 10, and then it jumps and rises linearly thereafter. But this is just an example. What one would do in an application, is to construct a realistic loss function, estimated based on potential damage to property, or individuals, or infrastructure, which we would then combine with the fitted probability model, using the formulae that I gave you earlier. So for each outcome there's going to be a probability and a potential loss, which we combine to calculate the overall expected risk.

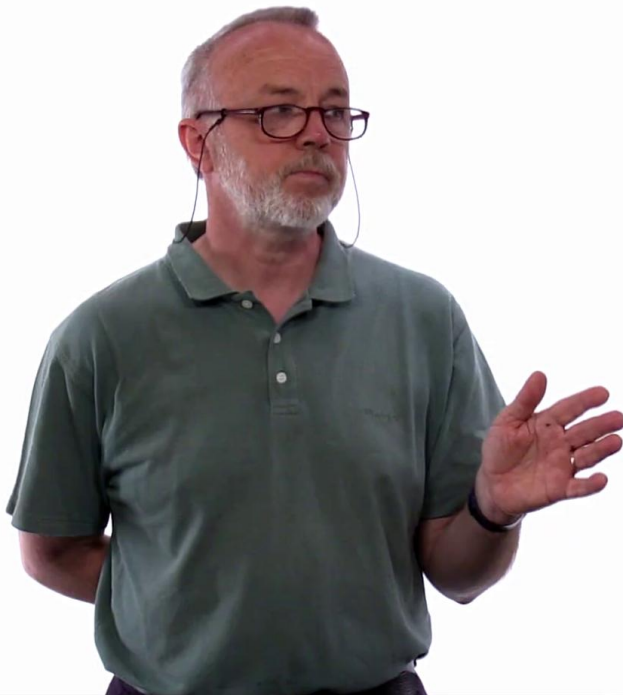
Notes

Summary



7m 27s

Variation and uncertainty



- Useful to distinguish variation and uncertainty
 - Variation is present because, even if we know the probability model, the outcome is not pre-determined
 - Uncertainty is present because in any application the probability model must be estimated from limited data
- In applications, both variation and uncertainty must be quantified, so far as possible

When thinking about a statistical problem like this, it's useful to distinguish between two different ideas, the ideas of variation and uncertainty. There's always variation present in a probability model, because even if we knew the model perfectly, which of course we don't, the outcome is not pre-determined. There's no deterministic mechanism to tell us what the rainfall will be on any given day. Uncertainty is a different concept. It's going to be present, because in any application we're going to have to estimate our probability model from limited data, and therefore we're going to be unsure about whether the probability model is really a good one, whether it's an adequate representation of the data. And in applications, we need to quantify both the variation coming from the underlying probability model, and the uncertainty that arises because we're not sure what the probability model is, or what it should be, as much as possible.

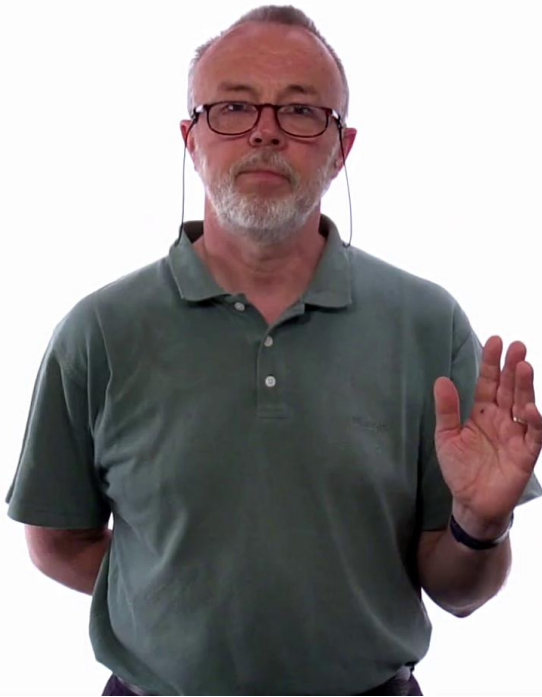
Notes

Summary



8m 49s

Main points



- Represent data by a fitted probability model, which we can use to estimate probabilities of unseen outcomes
- Estimate risk by combining the model probabilities for outcomes with their estimated losses

A Resilient Future: Science and Technology for Disaster Risk Reduction

So to go over the main points again, what we do when we try and estimate risk, is to estimate our data by an underlying probability model, which we're then going to use to estimate the probabilities of different potential outcomes. We estimate risk by combining the model probabilities for the different outcomes, with the corresponding estimated losses.

Notes

Summary



9m 52s

Main points



- Limitations:
 - The data are always limited, and are often compromised
 - The probability model rests on assumptions that may not be exactly true, and whose ramifications may be unclear
 - The losses are estimated
 - There may be black swan events that could not have been anticipated
- Nevertheless, we must do our best!

So there are some important limitations of what I've just described. The data are always limited, and are often compromised in some way. We may not be sure how reliable they are. The probability model will rest on assumptions that may not be exactly true, and whose consequences may be unclear. And of course, the losses are estimated. And then finally, there may be so called "black swan" events, that one could not have anticipated within the given model. Nevertheless, we have to do our best, as scientists and engineers, to try and estimate risks, because accurate estimation is key to mitigating them subsequently.

Notes

Summary



10m 17s

Image credits

Image credits in order of appearance (please consult the annex to access the links):

- Cover Picture: “dice” by [Ciarán Archer](#) is licensed under [CC BY 2.0](#) (8) EPFL – Prof. Anthony Davison
- (1) Modified from “[South America Continent Location Map](#)” by Blatant World is licensed under [CC BY 2.0](#) (9) EPFL – Prof. Anthony Davison
- (2) Modified from “[Vargas in Venezuela](#)” by TUBS is licensed under [CC BY-SA 3.0](#)
- (3) Modified from “[Vargas in Venezuela](#)” by TUBS is licensed under [CC BY-SA 3.0](#)
- (4) “[Montañasdeslave](#)” by David Virgilio Rodrigues Silva is licensed under public domain
- (5) EPFL – Prof. Anthony Davison
- (6) “[Caraballeda 1999 Deposits and Damage](#)” by USGS is licensed under public domain
- (7) EPFL – Prof. Anthony Davison

Thank you.

Notes

Summary

10m 59s

