Figure 7.2. NUMERICAL PROBABILITIES: Some Examples

EM9114: The Globe and Mail, December 7, 1991, page D10

MIND AND MATTER

STEPHEN STRAUSS

Relax Chicken Little. It's statistics, not the sky, that we should be questioning

TODAY I run the gamut of science from A to A – that is, from AIDS to asteroids.

What links two such seemingly disparate subjects is the way in which they spew up an iconic numerology in their wake. By iconic I mean that once a figure, suitably cosseted in qualifications, appears in print, everyone and his dog pick up on the number and make it gospel.

Consider the thesis that over the next 50 years North Americans have a one-in-6,000 risk of being killed by an asteroid. In June, The New York Times published the fright/scare/risk number and it immediately began appearing everywhere – including, in somewhat different forms, twice in this newspaper.

It's terrifying because it projects death by asteroid as a much greater risk than from tornadoes (one in 50,000), airplane crashes (one in 20,000) and just a bit lower than by electrocution (one in 5,000).

Of a population of about 280 million people in the U.S. and Canada, it seems that over a span of 50 years about 5,600 will be killed by tornadoes, 14,000 by airplane crashes – but more than 46,000 will expire after being bonked by space rocks.

Cover your heads, Chicken Littles.

Or should you?

"Ask what's the most likely number of people who will be killed by asteroids in the next two centuries and the answer is zero," admits Clark Chapman, a planetary scientist in Tucson, Ariz., who with his colleague David Morrison originated the Times' figures.

What the pair worked from are estimates that a large asteroid (about 1.25 kilometres across) has struck Earth on average every 300,000 years. They then factored in a hypothetical death toll from the collision.

The assumption is that it isn't being struck on the head that will kill people but the collateral effects – tidal waves and planet-cooling dust clouds. For their analysis, the scientists assumed that every living human on the face of the planet met his or her maker on asteroid-impact day. Thus, their one-in-6,000 figure arises simply by dividing 50 years into a 300,000-year possible arrival time and assuming total destruction.

'Ask what's the most likely number of people who will be killed by asteroids in the next two centuries and the answer is zero'

It sounds so neat, but sit on the numbers just a little and they squish all over the place. There has been no discernible cycle to asteroid collisions.

"You might not have anything that size for five million years and then three in a day," comments Richard Grieve, a geophysicist with the Department of Energy, Mines and Resources in Ottawa.

Equally murky is what a real death toll might be. Various estimates included in a table of a paper Dr. Chapman presented in October show that while some have predicted a total kill-off, the scientific consensus forecasts somewhere between 10 and 20 million fatalities worldwide. If North America got its roughly one-million share then the risk of being killed by an asteroid in the next 50 years shrinks to one-in-1.6 million. This is, on a risk-analysis basis, less than the likelihood of being killed by fireworks.

All of which leads me to doubt the wisdom of using any iconic asteroid number. All that can be sensibly said about the risk of the sky falling on us is this: If a rock, the diameter of two CN Towers laid end to end, does hit the earth, a lot of you are likely to die. But in the meantime, all potential Chicken Littles should throw out their moldy salmon, stop smoking and fasten those seat belts.

And now to AIDS. One of the questions continually asked after the basketball star Magic Johnson was diagnosed as being infected with the HIV virus was when Magic or anyone else in his position will come down with AIDS proper. The most common, but unfortunately wrong, answer was 10 years.

What a variety of studies do suggest is that somewhere between 10 and 11 years after they were infected, *half* the people observed had full-blown cases of AIDS. When will all HIV-positive people get AIDS? Maybe very much later, and maybe, just maybe, never.

The cause for some optimism is an ongoing study of 524 gay men in San Francisco. The study began in 1978 to document the spread of the sexually transmitted hepatitis B virus: AIDS researchers have been able to make use of blood samples which were frozen after the original hepatitis analysis.

This has allowed them to identify men who were or became HIV positive in the period 1978-191. As might be suspected, most – 361 – men have come down with full-blown AIDS. Another 116 show some signs of immune damage associated with the syndrome. But 47 men, 9 per cent of the original group, are perfectly healthy.

And of that number, 31 men have taken none of the anti-viral drugs that are given to HIV-positive people to ward off the onset of the disease. Their bodies, for reasons that the San Francisco researchers suspect may be linked to a genetic predisposition, are just better at fighting off AIDS.

"We don't know how long they will take to progress to AIDS or even if they will progress," says Susan Buchbinder, a doctor with the San Francisco Department of Public Health, which is conducting the study.

I hope the latter is so because this is a disease where even a scintilla of statistical hope looms large over a field so littered with the dead and dying.

1 In the fifth paragraph of the left-hand column of the article EM9114 reprinted above, expected numbers of deaths in the next 50 years in the U.S. and Canada due to tornadoes, aeroplane crashes and asteroids are given. Using information in the article, explain briefly how these numbers are obtained.

- How many deaths would be expected due to electrocution? Show your reasoning.
- Explain briefly what is meant by the word *expected* as it is applied to the numbers of deaths from the four causes.
- In the third paragraph of the right-hand column of the article, Stephen Strauss gives a reduced risk of death due to asteroids as one-in-1.6 million. Show how this modified figure is obtained.

(*continued overleaf*)

2 In the eighth paragraph of the left-hand column of the article EM9114 reprinted overleaf on page 7.7, Dr. Clark Chapman is quoted as saying: *Ask what's the most likely number of people who will be killed by asteroids in the next two centuries and the answer is zero.*. Indicate briefly what can be inferred from this statement about the *shape* of the distribution of the number of deaths due to asteroids.

3 In the second paragraph of the right-hand column of the article EM9114 reprinted overleaf on page 7.7, Richard Grieve is quoted as saying: *You might not have anything that size for five million years and then three in a day.* Discuss briefly the probabilistic issue(s) raised by this statement.

4 The sixth paragraph of the right-hand column of the article EM9114 reprinted overleaf on page 7.7 mentions attributes of the population of HIV-positive individuals; for example, the distribution of their times from HIV infection to having "full blown" AIDS. What two *other* attributes are mentioned? Explain briefly.
- What, if any, is the *relationship* between these two attributes? Explain briefly.
  - Indicate briefly the basis of Stephen Strauss' suggestion that some HIV-positive people may *never* get AIDS.

EM9013: The Globe and Mail, January 24, 1990, page C13

Professors see similarities to Ewoks in 1978 script

BY JIM MORRIS
Canadian Press

CALGARY

Two professors at the University of Calgary testified Monday there are striking similarities between the Ewok characters in George Lucas' film *Return of the Jedi* and the furry creatures of the same name in a script written by a Calgary man.

James Dugan, an actor and head of the university's drama department, said the characters are similar enough that they cannot have been created independently.

"The very similar general description and habits of the creatures seem to defy coincidence," he said.

Dugan was testifying at a lawsuit filed by Dean Preston, a Calgery writer who says he created the Ewok characters in 1978, five years before they appeared in Lucas' lucrative *Star Wars* trilogy.

Dugan said that if both Lucas and Preston were his pupils in a screenwritng class, he would question Lucas' integrity. "If I had received *Space Pets* from Preston as a final project in 1978 and *Return of the Jedi* in 1981, I would bring Lucas before the dean on a charge of plagiarism," he said.

English professor Janis Svilpis, who teaches science-fiction writing, also described the resemblance between Lucas' Ewoks and the bear-like creatures found in Preston's *Space Pets* script.

"Given that *Space Pets* and *Return of the Jedi* are closely related, and given that no common source for both is easily visible, it is somewhat more probable ... that *Space Pets* is the source text for *Return of the Jedi* rather than the reverse," said Svilpis.

Both men were presented as expert witnesses by A. Webster MacDonald, Preston's lawyer in the $150-million suit against Lucas. Both read from reports they prepared comparing Preston's script and Lucas' movie.

The reports noted the Ewoks and Preston's furry characters were similar in appearance, had primitive cultures, used spears and rocks as weapons, and lived in tree huts connected by vines and walkways.

Preston says he mailed the *Spapce Pets* script to Lucas but received no reply and the script was never returned.

Lucas says Ewoks evolved from the eight-foot, bellowing Wookie character in the original *Star Wars* script he wrote in 1974. Testifying last week, the bearded filmmaker also said his company has a strict policy of returning all unsolicited scripts unopened.

Lucas' lawyer, W. Graham Dutton, wrangled with Svilpis over the professor's calculations that there are only 1,274 chances in 100 trillion that both Preston and Lucas could imagine the same name for their creation.

"You're being silly," Dutton said of Svilpis's calculation.

The derivation of the word *Ewok* has been a central issue in the lawsuit.

Preston says the name was born while he was playing with phonetics and the phrase "he walks." Lucas testified the name was derived from switching around the letters of the word *Wookie* and his musings about the Miwok (pronounced Meewok) Indian tribe of northern California.

Dugan refused to accept Lucas' testimony.

"I don't get that as a short step," he told Dutton. "It doesn't sound to me that obvious."

Named in the suit are Lucas, Lucasfilm Ltd. and 20th Century Fox Canada Ltd.

5 In light of dates elsewhere in the article EM9013 reprinted above, comment briefly, from a *statistical* perspective, on the phrase ... *it is somewhat more probable* ... in the statement from Prof. Svilpis quoted in the third paragraph in the middle column of the article.

6 Can you suggest reasoning that leads to the probability, given in the second paragraph of the right-hand column of the article EM9013 reprinted above, of 1,274 chances in 100 trillion that both Prestom and Lucas could imagine the same name for their creation? It may or may not be of assistance to note that 1,274 is the product $2\times7\times7\times13$ and 100 trillion is (presumably) 10^{14}.

EM9207: Environment Canada Fact Sheet: Atmospheric Environment Service, May, 1992
Catalogue No. En37-24/3-1992E, ISSN 0715-0040, ISBN 0-662-19732-1

Apart from addressing a topic of wide interest, the article EM9207 reprinted (without its four illustrations) on the facing page 7.9 provides useful insights into both weather forecasts and, of more immediate importance, the (unfamiliar and often difficult) concept of 'probability.' For instance: (continued on page 7.10)

(*continued*)

Figure 7.2. NUMERICAL PROBABILITIES: Some Examples (continued 1)

PROBABILITY OF PRECIPITATION

Improving the weather forecast with the aid of numbers

Weather is an incredibly complex phenomenon. Despite the use of computers, satellites and skilled forecasters, it is difficult to predict future precipitation. Rain or snow cannot always be forecast with a simple yes or no. However, the probability of precipitation helps to gain a better understanding of what the future weather holds. It offers us another way to view the precipitation forecast.

To a large extent, Environment Canada's weather service has always been in the probability business. Forecasters have, at times, qualified their predictions with subjective phrases like "scattered showers" or "possibility of snow tomorrow." Now, forecasters are expressing this type of probability by numbers, adding a probability of precipitation statement to regional forecasts.

What is probability of precipitation?

Probability forecasts are a subjective numerical estimate of your chances of encountering measurable precipitation at some time during the forecast period. For example, a 40% probability of rain today means there are 4 chances in 10 of your getting wet today.

Basically, probability of precipitation forecasts allow the forecaster to express his/her degree of belief in the likelihood of precipitation occurring using percentages. In the regular forecast he/she might say "one or two showers today," but in the probability of precipitation he/she might predict 20 percent probability of precipitation. Again, if the public forecast says "there is a good chance of precipitation tomorrow," the forecaster might follow this by adding: 80 percent probability of precipitation.

Probability of precipitation and the weather forecast

ISOLATED SHOWERS THIS MORNING. CLOUDY THIS AFTERNOON. A FEW SHOWERS TONIGHT. RAIN TOMORROW.

PROBABILITY OF PRECIPITATION 30 PERCENT TODAY, 60 PERCENT TONIGHT, 90 PERCENT TOMORROW.

As the above example shows, probability forecasts are part of the regional forecasts for up to three specific periods: today, tonight and tomorrow. Today refers to the time period 6 am to 6 pm; tonight refers to the time period 6 pm to 6 am and tomorrow refers to the time period 6 am to midnight the following day.

Probabilities are given in 10% increments ranging from 0-100%. However, most forecasts tend to stay within the 10-90% range since forecasts cannot always be certain that precipitation will not or will occur due to the very complex nature of the atmosphere.

How is the probability forecast determined for a particular region? The forecaster studies the current weather situation, including wind and moisture patterns and determines how these patters will change with time. Factored into this is the effect of terrain, long-term weather statistics and the character of the precipitation (showery or continuous). In the case of showers, it is unlikely that precipitation will occur at all points in the region, whereas continuous rain is likely to be more widespread.

The procedures for producing a probability forecast do not radically differ from those employed in producing a regular forecast, but the forecaster does spend more time checking more points in a region before determining the possibility of precipitation.

The limitations of probability forecasts

Probability forecasts cannot be used to predict when, where or how much precipitation will occur. For example, a 60% probability of snow today does not mean it will snow during 60% of the day. However, the probability figure does mean that there is a 60% chance of a measurable amount of snow falling at that location.

A ROUGH AND READY USER'S GUIDE TO PROBABILITY OF PRECIPITATION

Probabilities	
0%	No precipitation even though it may be cloudy.
10%	Dry weather with only one chance in ten of snow or rain falling.
20%	Dry weather still expected.
30%	Go ahead with your picnic, boating or ski plans but you may have to take shelter.
40%	An umbrella is recommended. Make alternate plans for outdoor activities that are susceptible to rain. Not a good day to pave the driveway. Keep your fingers crossed!
50%	It's even Steven on whether it snows or not. Be prepared for all eventualities.
60%	Want to water your lawn? The odds are favourable that Mother Nature might give you some help.
70%	Suggest cancellation of outside events. The chances for dry weather have shrunk to three in ten.
80%	Wet weather likely. Make appropriate plans.
90%	The occurrence of precipitation is a near certainty. Venture out if you enjoy walking in the rain or playing in the snow.
100%	Precipitation is a certainty.

A person who did not encounter precipitation during the period would be tempted to say that forecast probability should have been zero, while the individual who did have precipitation during the same period would say the probability should have been 100%. Statistically, one cannot determine the accuracy of a single probability forecast. The accuracy can only be verified after a number of forecasts. A 30% probability of precipitation forecast is accurate if the same forecast was made on one hundred occasions and if it rained on 30 of those occasions.

The weather office forecasts a probability of precipitation for your region based on the information available at the time the forecast is issued. As the time of the event nears and more information becomes available, the forecaster is more accurate in predicting the probability of precipitation.

It should be emphasized that probability forecasts are not meant to replace regular weather forecasts. They simply supply a useful numerical addition to the existing forecasts.

The utility of probability forecasts

Probability forecasts are more than a novelty or an interesting variation on existing forecasts. They have considerable practical value and should allow many people to make more rational decisions about weather sensitive activities.

For example, a contractor might use probability of precipitation forecasts to decide whether or not to cure a concrete pad. He/she might juggle a 40 percent probability of rain forecast with such factors as actual dollor loss (due to rain) and expected loss based on the contractor's probable loss over time if the probability is accurate.

Probability forecasts are also useful to farmers making crop decisions, to promoters organizing outdoor sporting events and to your own personal picnics or skiing trips.

Generally, probability of precipitation forecasts aim to establish better communictions between the weather service and the public. Adding numbers to supplement the existing forecasts provides Canadians with more useful weather forecasts.

(*continued overleaf*)

In the second paragraph in the right-hand column of the article EM9207 reprinted overleaf on page 7.9:

- *Statistically, one cannot determine the accuracy of a single probability forecast.*
- *The accuracy can only be verified after a number of forecasts. A 30% probability of precipitation forecast is accurate if the same forecast was made on one hundred occasions and if it rained on 30 of those occasions.*

These statements remind us of the key statistical idea of *behaviour under repetition,* introduced in Figure 1.1 in Part 1 of these STAT 220 course Materials

A related idea arises with making *one* measurement of an unknown value – it provides *no* information about measurement *error,* as discussed in more detail on the upper half of page 5.60 in Appendix 5 of Figure 5.7 in the STAT 231 Course Materials; *repeated* measurements of the *same* variate can quantify measuring *imprecision* and, in conjunction with a *standard, inaccuracy.*

- Measuring an *un*known value is, of course, to be distinguished from measuring a (known) standard to *calibrate* a measuring process.

In the third paragraph in the right-hand column:

- *As the time of the event nears and more information becomes available, the forecaster is more accurate in predicting the probability of precipitation.*

This statements reminds us of the relationship between uncertainty and completeness of information and, hence, of limitations imposed on Answers by incomplete information; these matters were also introduced in Figure 1.1.

In the fourth paragraph in the left-hand column:

- *Basically, probability of precipitation forecasts allow the forecaster to express his/her degree of belief in the likelihood of precipitation occurring using percentages.*

This statement reminds us that *personal probabilities* are little more than subjective assessments of a person's *strength of belief,* the latter being a more evocative description and free from the implication that such assessments have defined mathmatical properties like 'real' probabilities (as discussed in, for example, the following Figures 7.4, 7.5 and 7.8).

In the eighth and tenth (the second-last and last) paragraphs in the left-hand column:

- *In the case of showers, it is unlikely that precipitation will occur at all points in the region, whereas continuous rain is likely to be more widespread.*
- *Probability forecasts cannot be used to predict when, where or how much precipitation will occur.*

These statements provide useful information for users of weather forecasts.

1995-04-20