The “100-year flood”

Douglas A. Kerr

Issue 1
January 20, 2022

ABSTRACT

We might read that some community, as a result of a severe series of rainstorms, was hit by a “100-year flood”. That sounds like a pretty bad one. But just what does that expression mean?

In this article, I describe what is the usual formal definition behind that description, and discuss some if its implications. An appendix gives an equation used in the related calculations.

INTRODUCTION

We might read that some community, as a result of a severe series of rainstorms, was hit by a “100-year flood”. That sounds like a pretty bad flood. But just what does that expression mean?

If someone has no real background in statistics, he might explain that to a friend as meaning “the kind of flood that we could expect to only hit every 100 years”. But, especially if we have some knowledge of statistics (perhaps at the high school AP level), we might wonder what **that** means.

THE ACTUAL DEFINITION

In fact, if we probe a little deeply, we find that as used by the USGS, when we read that, for the magnitude of flood that is usually described in the popular press as a “100-year flood”, their real technical definition is this (in my words):

a. A flood of such magnitude that, over a long period of time, the average interval between floods of essentially that magnitude occurring is 100 years.

Here, 100 years is said to be the *(mean) recurrence interval* of floods of at least the magnitude of interest.

This definition is precisely equivalent:

b. A flood of such magnitude that, over a long period of time, the average rate of occurrence of floods of essentially that magnitude is 0.01 (1/100) per year.

Here, 0.01 per year is said to be the *(mean) occurrence rate* of floods of the magnitude of interest.

Either of those definitions depend on a mathematical fiction: that the "flood situation" is unchanged over that "long period of time". If we think of that period of time as being perhaps several thousand years, that is unlikely to actually be so. But we need to imagine such a thing for the definitions to make sense.

Commonly the "real" definition is stated as:

c. A flood of such magnitude such that the probability of one of essentially that magnitude occurring in any given year is 0.01 (1%).

That is not precisely the same thing as the definitions in (a) and (b), but for a recurrence interval like the one in the example, it would lead to almost the same statistical ramifications (as we will see shortly).

In any case, flood scientists deprecate the use of the term "100-year flood" in favor of one of the descriptions given above as (a) or (b).

THE POISSON PROCESS

That definition is basically rooted on the occurrence of floods of any given magnitude being what is known in the field of statistics as a "Poisson process". The probability distribution of events in that process (usually in time) is called the "Poisson distribution". Its defining features are:

- The average rate of an event occurring (per unit time) is constant over time. (At least over the period in which we are interested.)
- The probability an event will occur in an given interval of time is not affected by the past history of the event (for example, if "one just occurred").
- Two events cannot occur at the same time.

PROBABILITY OF OCCURRENCE

The probability that exactly one of the defined events will occur in a certain stated time period (or, if we prefer, the probability that at least one will occur), or, for example, the probability that exactly two will occur, or the probability that at least two will occur, can be calculated directly. Consider that we are again speaking of floods, in particular of such a magnitude that they are defined as "100-year" floods under the formal definitions above.

Now we ask, with regard to any arbitrary period of one year, what is the probability that exactly one such flood will occur? That calculates to be 0.0099, just a wee bit short of the 0.001 we might have expected.

The formula used to determine this is given in Appendix A.

We might have thought that, since the average rate of occurrence of such a flood is 0.01 per year, then the probability of one occurring in a one-year period would be 0.01. But it doesn’t work quite that way.

While we’re in the neighborhood, we ask what is the probability that, in any arbitrary period of one year, that **one or more** such floods will occur. That is 0.00995, just a wee bit greater than for “exactly one occurrence”. That is because the probabilities of having, 2, then 3, then 4, etc. occurrences (all of which would be included in the “one or more” probability) are very small.

Now we will shift to any arbitrary period of 10 years. There, the probability of having one or more such floods is about 0.095, noticeably less than the 0.1 we might have expected.

We might have thought that, since the average rate of occurrence of such a flood is 0.01 per year, then the probability of one occurring in a 10-year period would be 10 times as great, or 0.1. But, as we see again, it doesn’t work that way.

Finally we will shift to an arbitrary period of 100 years, the namesake of our flood type. There, the probability of having one or more such floods in such a period is about 0.632. I hesitate in this case to say “what we might have expected”. The probability of having two or more such floods in any given 100-year period is about 0.264; the probability of having three or more is about 0.080.

WHAT IF WE JUST HAD ONE?

What about this outlook: “If we just had a ‘100-year flood’, surely it will be about 100 years before we have another. Sadly, it doesn’t work that way. If the basic statistical properties I have suggested above actually hold, then if we had such a flood this year, the probability we will have one next year is still very nearly 0.01 (1%).

TOO OFTEN?

In that same vein, we may hear that the town of Fairview has just had a “100-year” flood, and in fact it had also had a “100-year” flood only 30 years ago. Is that anomalous?

It is not anomalous. The nature of the Poisson distribution is such that a scenario like this can exist, consistent with the Poisson model. As we saw above, the probability that there will be at least two “100-year” floods in a given 100-year period is about 0.264.

IN REALITY

If we actually modeled the very complicated mechanism by which such a flood might occur in some particular town, we would recognize that the Poisson model does not actually represent reality, even in the short term. For example, if the principal mechanism of such a flood is the overtopping of the Green Lake Dam, the occurrence of a flood may erode the dam threshold, thus raising the probability that a subsequent flood will occur in any given time period (contrary to one of the “requirements” for application of the Poisson model). Thus the handy relationships I discuss in this article would not really result.

That having been said, in any case, when we hear about Green Lake City having suffered a “100 year flood”, we know to be cautious in thinking about what that means.

-#-

Appendix A

Formula

The fundamental formula for, under the Poisson process model, calculating the probability that there will be a certain number of events in a certain interval is:

$$P_k = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (1)$$

where P_k is the probability that there will be k events in that certain interval and λ is the *mean event rate* in events per the duration of that certain interval. The factor $k!$ is the factorial of k, the product of all integers from 1 up through k. Note that k must be an integer; λ need not be.

Suppose we know that the average number of customers that, during the one-hour period 2:00-3:00, enter a bakery is 2.5 (and that the behavior of this flow is Poissonian). (That is a rate of 2.5 customers per hour.)

Then, if we wish to know the probability that 4 customers will enter the bakery during any one-hour period (at that time of day), we can use formula 1, with λ=2.5 and $k=4$. The result is 0.13.

What does that mean? Well, it means that over a long time (assuming the nature of the customer flow at that time of the day remains unchanged), we could expect (n the statistical sense) to have exactly 4 customers during that hour in about one day out of 8 (approximately 1/0.13).

What is the probability that, in some certain one-hour period (at that time of day) there will be no customers? Here we use λ=2.5 and k-0 (and note that 0! is defined as 1). The result is 0.082.

-#-