When Is It Time to Worry About Clustering?

A Poisson distribution tutorial for SH&E professionals

By Pat L. Clemens

CAN RANDOMNESS PRODUCE loss event "clusters?" For example, suppose in a large workforce of individuals performing nearly identical tasks under identical circumstances, one or only a few workers suffer an apparently inordinate number of injuries. Or, suppose a small number of identical mechanical systems in a large fleet appear to be responsible for an excessive number of failures. Are clusters of such events results of haphazard chance or something more insidious? The Poisson distribution provides an easy-to-use engineering tool for diagnosing these cases. Results help pinpoint which instances can be explained by randomness and which may deserve engineering or management attention.

Background: A Perplexing Problem

Consider a nationwide baked goods company that has a fleet of 82 identical dough-mixing machines in widely scattered plants. The machines are all the same age and their production rates are well matched. They experience similar service stresses, rates of use and maintenance routines. Yet, over a five-year period of service, during which 33 of the machines have experienced no failures, four machines have been responsible for nearly 20 percent of 69 machine failures experienced by the fleet. Four faulty machines or statistical mischance?

Next, consider an aircraft manufacturer with production workers engaged in metal trimming operations. Of 72 workers performing similar work, one has suffered nearly 12 percent of all hand cuts over an extended observation period—despite the fact that all have equal protection and have received the same training. Carelessness or a probabilistic expectation?

The "clustering" of loss outcomes such as these is vexing to the SH&E professional, whether it occurs in occupational safety or in system safety practice. The SH&E professional is often tasked with helping management decide whether to retrain or transfer the possibly careless worker, or to replace the suspect system. These decisions are especially troubling because one must find a reason for confidence that the clustering is not simply an effect of statistical randomness but, rather, is a consequence of a vulnerable system (in the first example) or a heedless worker (in the second).

A Useful Diagnostic Aid

The Poisson exponential distribution is an easily mastered analytical approach for use by the SH&E professional seeking confidence that a cluster of loss events is not simply a product of chance but may have a more insidious underlying cause that deserves attention (Ash; Green; Kumamoto and Henley; Raheja; Roland and Moriarty). Poisson modeling is widely employed as an analytical tool in reliability engineering, is often found in quantitative system safety analyses, and is frequently used in epidemiological medical studies (Hanley and Lippman-Hand), but it is rarely applied in occupational safety practice—where its use can also be beneficial.

The Poisson distribution models the probabilities that clusters of events will be experienced by individual items within a population. It assumes that event occurrences are governed only by randomness. In this context, events of concern are loss events. Such events may be employee injuries or system failures, for example. Experience-based field data containing apparent clusters can be compared with results of the Poisson analysis for insight into the likelihood that clusters of actual loss events may have arisen by chance.

Figure 1 presents the equation for the Poisson distribution and defines key terms. Here, "system" represents the individual worker or piece of equipment—anything to which the analysis may be applied. "Losses" refers to injuries or system failures; it is their probabilities of occurring in clusters that is to be explored.

An Equipment Example

Let's return to the example of the 82 identical dough-mix-

Pat L. Clemens, P.E., CSP, *performs system safety engineering work for Sverdrup Technology Inc., Tullahoma, TN. A past president of the Board of Certified Safety Professionals, he has developed and implemented system safety programs in both government contracting and in the private sector. Clemens teaches system-safety-related courses for various private corporations, NASA and ASSE, and on university campuses. He is a professional member of ASSE's Middle Tennessee Chapter.*

Figure 1

The Poisson Distribution

$$P_{(n)} = \frac{(m)^n \epsilon^{-m}}{n!}$$

$P_{(n)}$ = probability, per system, of n losses
n = loss count examined; i.e., cluster size
ϵ = Napierian base (2.718 . . .)
m = long-term average loss rate, per system

ing machines that over a five-year period of exposure experienced a total of 69 failures. Failures were distributed among the machines (Table 1). In this example, 33 machines experienced no failures; 34 had one failure each; two failures were suffered by each of 11 machines, etc. These numbers of failures are the cluster sizes and become the *n* terms to be used in successive solutions of the Poisson expression in Figure 1.

To evaluate the *m* term of Figure 1, one must recognize that 69 failures occurred among the 82 machines over the five-year period:

$$m = \frac{69}{82} = 0.841 \frac{\text{failures}}{\text{machine}}$$

The n! (n factorial) in the equation signifies:

$$n! = 1 \times 2 \times 3 \times \cdots (n\text{-}1) \times n$$

Thus, to explore the probability of a cluster size of four (i.e., n = 4):

$$n! = 1 \times 2 \times 3 \times 4 = 24$$

Using these values of *m* and *n*, one can find $P_{(n)}$—the probability that any one machine among the 82 will suffer a cluster of four failures during the five-year exposure interval:

$$P_{(n)} = \frac{(m)^n \epsilon^{-m}}{n!}$$

$$P_{(4)} = \frac{(0.841)^4 \times 2.718^{-0.841}}{24}$$

$$P_{(4)} = 0.00899 \text{ per system}$$

Rounding this result, it is interpreted to mean that any single machine selected randomly from the fleet has a probability no greater than 0.009 (0.9 percent) of failing four times in the five-year period. This seems a rather low probability in view of the operating experience. However, the fleet has 82 independent machines, each with this same probability of experiencing four failures. Thus, one must multiply $P_{(n)}$ by that population to approximate the number of machines expected to suffer four failures in the fleet: 82 x 0.00899 = 0.737 ≈ 0.74

Since failures occur only as whole numbers, this result can reasonably be rounded to 1.0. In fact, as Table 1 shows, one machine did experience four failures during the five-year period. Unfortunately, at the outset of the five-year period, it would be impossible to know which of the 82 machines would suffer those four failures as all would have had equal likelihood.

The Poisson results for clusters of failures from zero through four have been computed using this approach (Table 2). As this shows—perhaps surprisingly—the numbers of failures to be expected based on the Poisson analysis are rather closely matched by actual experience of the fleet. Were the population of machines larger or the period of observation longer, an even closer match would be expected.

Table 1

The Mixing Machine Failure Record

Machines	Failures*
33	0
34	1
11	2
3	3
1	4

**Cluster size, n*

Table 2

Expected vs. Actual Failure Distribution

Failure Count (n)	$P_{(n)}$	Expected Failures*	Actual Machine Failure Experience†
0	0.431	35.4	33
1	0.363	29.7	34
2	0.153	12.5	11
3	0.043	3.51	3
4	0.009	0.74	1

$P_{(n)}$ = probability per machine of n failures
**82 x $P_{(n)}$ per machine*
†Number of machines suffering n failures over five-year period (from Table 1)

An Occupational Safety Example

Now, translate this example into the realm of occupational safety and health. Replace the dough-mixing machines with the 72 metal trimmers and substitute their accumulated 51 workplace hand cut injuries (which required either first aid or greater care) for the machine failures. Table 3 presents the injury distribution.

How might the worker who experienced those six injuries in that period be viewed, when 34 others performing the same work under the same circumstances had none? Is that worker careless, a victim of randomness or should other factors be explored?

Again evaluating the *m* term of Figure 1:

$$m = \frac{51}{72} = 0.708 \frac{\text{hand cuts}}{\text{worker}}$$

Applying it in the Poisson expression (Figure 1) as before, arbitrarily using an example cluster size of 3:

$$P_{(3)} = \frac{(0.708)^3 \times 2.718^{-0.708}}{6}$$

$$P_{(3)} = 2.92 \times 10^{-2} \text{ per worker}$$

Thus, any selected worker has a probability of a just less than 0.03 (three percent) of suffering three hand cuts over the study period. Because this population includes 72 workers, randomness dictates that

Table 3
The Hand Cut Injury Record

Workers	Injuries*
34	0
30	1
6	2
1	3
0	4
0	5
1	6

Cluster size, n

Table 4
Expected vs. Actual Injury Distribution

Injury Count (n)	$P_{(n)}$	Expected Injuries*	Actual Worker Injury Experience†
0	0.492	35.5	34
1	0.349	25.1	30
2	0.124	8.89	6
3	2.92×10^{-2}	2.10	1
4	5.17×10^{-3}	0.372	0
5	7.32×10^{-4}	0.0527	0
6	8.64×10^{-5}	6.22×10^{-3}	1

$P_{(n)}$ = probability per worker of n injuries
**72 x $P_{(n)}$ per worker*
†Number of workers suffering n injuries over observation period

the number of them likely to suffer three injuries over the period is approximately $72 \times 0.0292 \approx 2.10$

Table 4 shows expected versus actual injury distribution in this example. Results show that for the most part the injury distribution actually experienced reasonably matches that predicted. An obvious exception is the one worker who experienced six hand cuts during the study period. Notice that the probability of this occurring in the 72-worker group is appreciably less than one percent.

Interpreting Results

Results of the analyses are of particular interest to the SH&E professional. In the fleet of 82 dough-mixing machines, randomness alone would lead to the expectation that a single machine would suffer four failures during the exposure interval. Similarly, it should be an expectation that just four of the 82 machines might experience 13 of the 69 failures experienced by the fleet. (One machine suffered four failures and three suffered three failures each for the total of 13.) The Poisson results assure the analyst that machine-to-machine differences alone cannot be held accountable for the evident clustering seen in the failure record.

Without the analysis, however, one might be tempted to condemn those four machines as failure-prone devices and to replace them. Doing so without making any other changes would give one no reason to expect better fleet performance during the next five-year exposure interval. During that next interval, it would not be surprising for another machine somewhere in the fleet to suffer four failures, and another group of three to experience three each, and so on.

In the case of the hand cut injury distribution among the 72 metal trimmers, the analytical results show that the apparent cluster of six injuries experienced by a single worker far exceeds the probabilistic expectation. Poisson modeling has provided assurance that pure randomness cannot account for this cluster. Other causes should be considered and remedies sought. All other cases in the injury count distribution fall at values that reasonably match expectations.

The analyses have also shown that if a company cannot tolerate having the loss event distributions and clustering shown in these examples, it must find ways to reduce the long-term average loss rate for the fleet or workgroup—which is the *m* term in the Poisson expression.

Conclusion

Before denouncing a piece of equipment or a particular worker for a rash of loss events, the prudent SH&E professional will confirm that the events are not readily explainable on the basis of random misfortune. Poisson distribution provides an uncomplicated and readily mastered method of verifying the clustering role played by chance—and may cause the SH&E professional to alter intuitive reactions to clusters. ■

Poisson Distribution Assumptions

In applying the Poisson distribution model as a means of exploring the likelihood that clusters of loss events are attributable to randomness, one must recognize the assumptions it imposes.

- It presumes that the items making up the population analyzed are identical, whether they are workers, subsystems, components, etc.
- It assumes that the items individually occupy either one or the other of only two states—injured or uninjured, functioning or faulty.
- A single, known event occurrence rate must be presumed to apply to all items in the population throughout the interval of exposure; in the first example presented, a long-term average of 0.841 failures per machine was used, based on the operating experience of the fleet of machines.
- The occurrence rate per item must be relatively small over the interval of study (i.e., <1.0). In the first example, it was 0.841.

References

Ash, C. *The Probability Tutoring Book.* New York: IEEE Press, 1993.

Green, A.E., ed. *High Risk Safety Technology.* New York: John Wiley & Sons, 1982.

Hanley, J.A. and A. Lippman-Hand. "Interpreting Zero Numerators." *Journal of the American Medical Assn.* 249(1983): 1743-1745.

Kumamoto, H. and E.J. Henley. *Probabilistic Risk Assessment and Management for Engineers and Scientists.* New York: IEEE Press, 1982.

Raheja, D. *Assurance Technologies.* New York: McGraw-Hill, 1991.

Roland, H.E. and B. Moriarty. *System Safety Engineering and Management.* 2nd ed. New York: Wiley-Interscience, 1990.

Your Feedback

Did you find this article interesting and useful? Circle the corresponding number on the reader service card.

RSC#	Feedback
36	Yes
37	Somewhat
38	No