the Availability Digest

www.availabilitydigest.com
@availabilitydig

Random Events Have No Memory

November 2014

A word of caution: This article is truly a Geek article. It is for those who enjoy heavy mathematics. However, having said that, the message of this article is important for everyone trying to understand a little about availability theory. Therefore, it begins with a general discussion that is for everyone.

If you consider yourself mathematically challenged, please read the introductory section, which will give you the important background that you need. The rest of the article is for math nuts and serves to prove the conjectures made in the first section.

Memoryless Variables

In availability, we talk about *mean time between failures (MTBF)* and *mean time to repair (MTR)*, where "mean" means" average." MTBF is the average time between system failures. MTR is the average time it takes to repair a failed system and return it to service (in some cases, we use *mean time to restore* for MTR, depending upon the context).

Availability (A) is the proportion of time that the system is operational – that is, it is the probability that the system will be up. Thus,

$$A = \frac{MTBF - MTR}{MTBF} = 1 - \frac{MTR}{MTBF}$$

Since the system must be either up or down, the probability that the system will be down - the *probability of failure (F)* – is

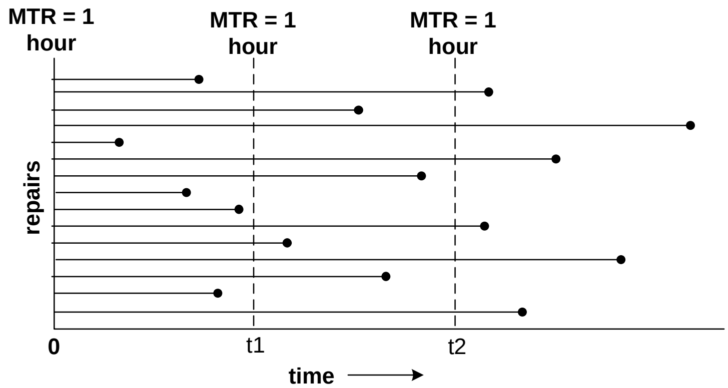
$$F = 1 - A = \frac{MTR}{MTBF}$$

MTBF and MTR are *random variables*. That is, they can take on any number of values. The system may fail in three months, and then it may not fail again for two years. It might take one hour to repair the system the first time, and 20 minutes to repair it the next time.

A key concept in availability theory is that MTBF and MTR are *memoryless* variables. That is, whether the system will go down or will be repaired in the next minute is absolutely independent of the past. If an operational system is observed, it may go down in the next minute or in the next year, regardless of its past failure history, so long as the average time between failures is MTBF.

Perhaps more troubling is that if a system under repair is observed, the time that it will take to repair from that time on will always average MTR, no matter how long the system has been under repair. For instance, if the MTR for a system is one hour, and the repair team has been working on the system for 45 minutes, the average time for the repair to be completed from that time on is still one hour.[1]

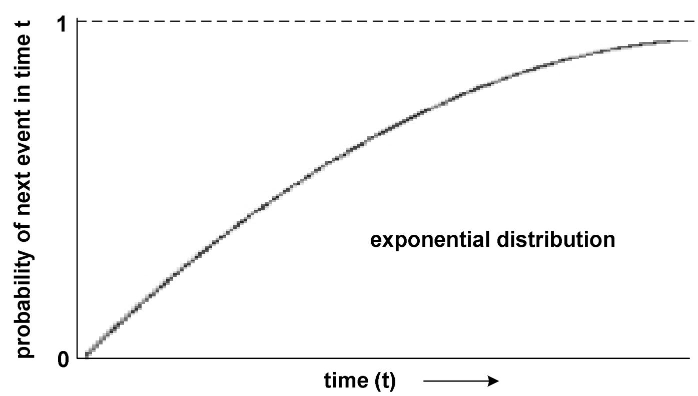
This can be envisioned from the following figure. Assume that we keep track of the repair times over several years and have an extensive database of repair times. We plot them as shown in the following figure, all starting at time = 0. If we measure the repair times from that point, we find that the average repair time is MTR. We then move out to a later time, t_1. By that time, some repairs have been completed. The average time to repair for the open repairs is still MTR. And so on.



This is the meaning of memoryless variables and is the basis for many analyses, not only availability theory but also queuing theory, for instance.

A more practical example may serve to illustrate the point. Consider your home telephone. You receive a call on it. What is the time from when you hang up until you receive the next call? It has nothing to do with the receipt of the first call. The next call may come immediately, or it may not come until the next day. The receipt of the next call has nothing to do with the prior history of calls. The time between calls is memoryless.

What is often not known is that there are two common probability distributions that describe random variables. One is the *exponential distribution* that gives the probability that an event (a failure or repair in our case) will occur in the next time *t*. For instance, what is the probability that a repair will be completed within the next hour?

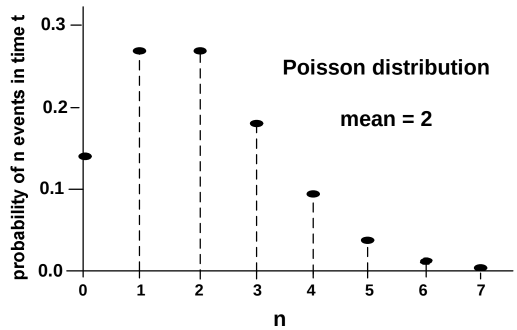


[1] See our article entitled Repair Strategies, *Availability Digest*; October 2014.
http://www.availabilitydigest.com/public_articles/0910/repair_strategies.pdf

$$\text{prob(repair} < t) = 1 - e^{-t/MTR} \qquad \text{exponential distribution}$$

The other probability distribution is the expected number of events, *n*, during an interval *t*. For instance, how many failures can we expect in the next five years? This distribution is known as the *Poisson distribution*:

$$\text{prob(n failures in time t)} = \frac{[t/(MTBF)]^n e^{-t(MTBF}}{n!} \qquad \text{Poisson distribution}$$



It is not obvious that random distributions, exponential distributions, and Poisson distributions characterize the same physical process. The rest of this article shows that the exponential and Poisson distributions are a direct result of the randomness of a physical process – that is, they describe memoryless random variables.

At this point, those who feel mathematically challenged may turn to other activities. The above has described the primary thrust of this article. [2]

The Poisson Distribution

The Poisson distribution provides the probabilities that exactly *n* events may happen in an interval *t*, provided that these events are independent. That the independence of events is the only assumption made is the reason that this distribution is so important.

The occurrence of an event is not at all dependent on what has occurred in the past, nor has it any influence on what will occur in the future. The process has no memory; it is memoryless. We call a process that creates such random events a *random process*. Note that randomness has to do with events: the failure of a system or its repair.

Let us determine the probability that exactly *n* random events will occur in a time *t*. We will represent this as $p_n(t)$:

$p_n(t)$ = the probability that *n* random events will occur in time *t*

The average rate of occurrence is a known parameter and is the only one we need to know. We will denote it by *r*:

r = average event occurrence rate

For availability analysis, *r* = 1 / *MTR* is the repair rate, and *r* = 1 / *MTBF* is the failure rate. Thus, on the average, *rt* events will occur in time *t*.

Since events are completely random, we know that we can pick a time interval, *t*, sufficiently small that the probability of two or more events occurring in that time interval can be ignored. We will note this

[2] The rest of this article was taken from W. H. Highleyman, Performance Analysis of Transaction Processing Systems, *Prentice-Hall*; 1989.

arbitrarily small time interval as Δt and will assume that either no events will occur or that one event will occur during this interval.

Let us now observe a process for a time t. At the end of this observation time, we count the number of events that have occurred. We then observe the process for Δt more time. The probability that one further event will occur in Δt is $r\Delta t$. The probability that no further events will occur is $(1 - r\Delta t)$.

The probability of observing n events in the time $(t + \Delta t)$ is

$$p_n(t + \Delta t) = p_n(t)(1 - r\Delta t) + p_{n-1}(t)r\Delta t \qquad n > 0 \ [(n-1) \text{ does not exist for } n = 0)] \qquad (1)$$

This equation notes that n events may occur in the interval $(t + r\Delta t)$ in one of two ways. Either n events have occurred in the interval t *and* no events have occurred in the subsequent interval Δt, or $(n - 1)$ events have occurred in the interval t *and* one more event has occurred in the subsequent interval Δt.

If no events occurred in the interval $(t + \Delta t)$, this relationship is written as

$$p_0(t + \Delta t) = p_0(t)(1 - r\Delta t) \qquad n = 0 \qquad (2)$$

That is, the probability of no events occurring is the probability that no events occurred in the interval t and that no events occurred in the interval Δt.

Equations (1) and (2) can be arranged as

$$\frac{p_n(t+\Delta t) - p_n(t)}{\Delta t} = -rp_n(t) + rp_{n-1}(t) \qquad n > 0 \qquad (3)$$

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -rp_0(t) \qquad n = 0 \qquad (4)$$

If we let Δt become smaller and smaller, this becomes the classical definition of the derivative of $p_n(t)$ with respect to t, $dp_n(t)/dt$. Denote the time derivative of $p_n(t)$ by $p'_n(t)$:

$$p'_n(t) = dp_n(t)/dt$$

We can express equations (3) and (4) as

$$p'_n(t) = -rp_n(t) + rp_{n-1}(t) \qquad n > 0 \qquad (5)$$

$$p'_o(t) = -rp_0(t) \qquad n = 0 \qquad (6)$$

Equations (5) and (6) are a set of differential equations. The solution to these equations is challenging even for math geeks and is given in the Appendix for the courageous. However, the result is

$$p_n(t) = \frac{(rt)^n e^{rt}}{n!} \qquad (7)$$

This is the Poisson probability distribution. It gives the probability that exactly n events will occur in a time interval t, given only that their occurrences are random with an average rate of r.

The Exponential Distribution

The exponential distribution deals with the probability distribution of the time between events. It can be derived from the Poisson distribution.

We assume again that events are occurring randomly at a rate of *r*. Let us consider the probability that, given that an event has just occurred, one or more events will occur in the following time interval, *t*. This is the probability that the time between events is less than *t*. If *T* is the time to the next event, we denote this probability as *p(T < t)*. From the Poisson distribution [Equation (7)],

$$p(T < t) = \sum_{n=1}^{\infty} \frac{(rt)^n e^{-rt}}{n!} \qquad (8)$$

That is, the probability that the next event will occur in a time interval less than *t* is the probability that one event will occur in time *t* plus the probability that two events will occur in time *t*, and so on.

Manipulating Equation (8), we have

$$p(T < t) = e^{-rt} \left[\sum_{n=0}^{\infty} \frac{(rt)^n}{n!} - 1 \right]$$

Since

$$\sum_{n=0}^{\infty} \frac{(rt)^n}{n!} e^{-rt} = 1$$

(the probability of some value of *n* occurring in time *t* is one), we can write

$$P(T < t) = 1 - e^{-rt} \qquad (9)$$

This is the exponential distribution. It gives the probability that the next event will occur in time *t*. Like the Poisson distribution, it depends only upon occurrences being random with an average rate of *r*.

Summary

It is commonly assumed in many areas of analysis that events occur randomly. The occurrence of a random event is independent of what has happened before, and the event has no impact on what will happen in the future. In availability theory, we assume that failures are random events as well as repair times.

Given a purely random occurrence of events, we can calculate the probability that a certain number of events will occur in a specific time interval. This is the Poisson distribution. We can also calculate the probability that the next event will occur by a specified time. This is the exponential distribution.

Memoryless events, the Poisson distribution, and the exponential distribution are all ways to characterize the same process – the random occurrence of events.

Appendix

Solving the Poisson Distribution Set of Differential Equations

We have noted that the Poisson distribution is given by the following system of differential equations:

$$p'_n(t) = -rp_n(t) + rp_{n-1}(t) \quad n > 0 \tag{5}$$

$$p'_n(t) = -rp_0(t) \quad n = 0 \tag{6}$$

The solution to this set of equations is due to T. L. Saaty.[3] Let us define a generating function *P(z,t)*:

$$P(z,t) = \sum_{n=0}^{\infty} z^n p_n(t) \tag{A-1}$$

If we should differentiate equation (A-1) *n* times with respect to *z*, we have

$$\frac{\partial^n P(z,t)}{\partial z^n} = n!p_n(t) + \frac{(n+1)!}{1!} zp_{n+1}(t) + \frac{(n+2)!}{2!} z^2 p_{n+2}(t) + ...$$

Setting *z* to zero, we obtain

$$\frac{\partial^n P(z,t)}{\partial z^n} = n!p_n(t), \quad z=0 \tag{A-2}$$

Thus, by differentiating the generation function *P(z,t)* *n* times with respect to *z*, setting z to zero, and dividing the result by *n!*, we obtain $p_n(t)$.

Let us now consider a time *t* and assume that *i* events have occurred up to time *t* = 0. That is, by the definition of $p_n(t)$,

$p_i(0) = 1$ probability that *i* events have occurred by time zero

$p_{n \neq i}(0) = 0$ probability that *n* events have occurred by time zero, where $n \neq i$

From Equation (A-1), for *t* = 0,

$$P(z,0) = z^i p_i(0) = z^i \tag{A-3}$$

Also, if *z* is set to 1, from Equation (A-1),

$$P(1,t) = \sum_{n=0}^{\infty} p_n(t) = 1$$ since the summation is over all values of *n*

Now let us multiply the differential-difference Equations (5) and (6) by z^n, obtaining

$$z^0 p'_0(t) = -rz^0 p_0(t) \quad n = 0$$

$$z^n p'_n(t) = -rz^n p_n(t) + rz^n p_{n-1}(t) \quad n > 0$$

[3] T. L. Saaty, Elements of queuing theory, *McGraw Hill*; 1961.

If we sum these over all *n*, recognizing that (*n* -1) does not exist for *n* = 0, we obtain

$$\sum_{n=0}^{\infty} z^n p'_n(t) = -r\sum_{n=0}^{\infty} z^n p_n(t) + r\sum_{n=1}^{\infty} z^n p_{n-1}(t) \qquad \text{(A-4)}$$

From Equation (A-1), the left-hand term of this equation is simply *∂P(z,t)/∂dt*. The first term on the right is –*rP*(*z*, *t*). The second term on the right is

$$rzp_0(t) + rz^2p_1(t) + rz^3p_2(t) + \ldots$$

$$= rz[p_0(t) + zp_1(t) + z^2p_2(t) + \ldots$$

$$= rzP(z,t)$$

Thus, Equation (A-4) can be written as the linear differential equation

$$\frac{\delta P(z,t)}{\partial t} = r(z-1)P(z,t) \qquad \text{(A-5)}$$

The solution to this is

$$P(z,t) = Ce^{r(z-1)t} \qquad \text{(A-6)}$$

which can be verified by substituting *P*(*z,t*) from Equation (A-6) into both sides of Equation (A-5).

The value of *C* is dependent upon how many items, *i*, are received by *t* = 0. Let us assume that at *t* = 0, zero events have occurred. In this way, $p_n(t)$ will truly be the probability of receiving *n* items in the subsequent interval *t*. From Equation (A-3), setting *i* = 0,

$$P(z,0) = z^i = 1$$

Thus, *C* = 1 in Equation (A-6) and

$$P(z,t) = e^{r(z-1)t} \qquad \text{(A-6)}$$

As we pointed out earlier with reference to Equation (A-2), $p_n(t)$ is derived from *P*(*z*,*t*) by differentiating *P*(*z*,*t*) *n* times with respect to *z*, setting *z* to 0, and dividing by *n*!. Performing these operations on Equation (A-6) yields

$$p_n(t) = \frac{(rt)^n}{n!}e^{-rt}$$

This is the Poisson distribution referenced in Equation (7).