www.simiode.org

SIMIODE A SYSTEMIC INITIATIVE FOR MODELING INVESTIGATIONS & OPPORTUNITIES WITH DIFFERENTIAL EQUATIONS

STUDENT VERSION
STOCHASTIC PROCESSES

Brian Winkel
Director SIMIODE
Cornwall NY USA

Abstract: We build the infinite set of first order differential equations for modeling a stochastic process, the so-called birth and death equations. We will only need to use integrating factor solution strategy or $\widehat{\mathrm{D}}$Solve in Mathematica for success. We work to build our model of random events which leads to the differential equations for $P_n(t)$, the probability that our system is in state n or the event, E_n, that the system is in state n, at time t. We compute the mean state value and then apply our newfound knowledge to solve some problems.

Keywords: random, stochastic, mean, Poisson process, traffic

Tags: data, model, infinite system, differential equation, industrial engineering, operations research, linear, separation of variables, integrating factors, mean, variance, parameter estimation, long, directed, develop model

STATEMENT

Introduction

Fasten your seat belts! We are going to take a fascinating journey into the world of random events and modeling uncertainty in which we will build a system of an infinite (yes you read that right!) number of differential equations to model random or stochastic events. We will use straightforward differential equations techniques to solve any and all differential equations that come our way in this study. We will apply this approach to model a number of practical phenomena.

These equations are studied and modeled in more detail in a very applied field called *operations research*. The method of our madness is to engage you in your own discovery and march on for results. *If you truly engage, ask questions, address issues we raise, then you will benefit. Otherwise, you will have done nothing more than read through stuff to the end and seen formulae which can be obtained in a much easier manner, e.g., look it up in some technical manual or formula set in* *a standard textbook in stochastic processes*. So the journey will be a rich one, filled with amazing twists and turns. Get your supplies together for the trip.

We shall take a journey into some basic probability notions, but firmly guide you through these introductory notions in context of building differential equations models. Basically, we are going to introduce you to a modeling tool to help you understand the notion of randomness and that tool is called a *stochastic process*. Then we will apply it to see if real phenomena are random by analyzing some real data.

Here we go . . .

Building a Model Conception of Random Events[5]

Think about what it means to be a random event? Consider some examples:

- Number of cars passing a fixed spot on an interstate in a given time interval?
- Number of channel switches per minute performed on the television by the resident "couch potato" in the dorm lounge.
- Number of particle emissions per second from a radioactive isotope.
- Number of active phone calls at a given three digit exchange.
- Number of hits to your Facebook page per hour during the day and night.
- Number of no-hitters per season in Major League Baseball.
- Number of people in the array of WalMart checkout counters.
- Number of information requests at the Amazon Customer Service Desk per hour.
- Number of points scored per minute during a college football game.
- Number of paramecia in a Petri dish each hour of a laboratory "run."
- Number of "reps" you can do in a bench press activity in the gym per 15 minute interval.

Let us consider one context, just as an illustration - but we will not build on in this scenario. This is just for ease in thinking about randomness. Consider, arrivals of telephone calls at a telephone switching station or messages at a university Internet port in a given interval of time. What might happen to a system over a short period of time, say $[t, t+h]$? "State" is a common word with a technical definition which we will see often in our study now. The word *state* means, the system is in state E_n at time t if exactly n messages have arrived in the interval $[0, t]$. Much of what is discussed here comes from traditional sources [3, pp. 255-301, 397-436.]. We seek to build a mathematical model for $P_n(t)$, the probability that the system is in state E_n at time t.

1) For several (pick ones that interest you) of the possible random events described in the list above define what the state, E_n, might be for each n and what a reasonable time frame would be over which to study this phenomenon. Explain what $P_n(t)$ means in each case for each value of $n = 0, 1, 2, 3 \ldots$ that you use. Attempt to sketch $P_n(t)$ for $n = 0, 1, 2,$ and 100 for each of the

possible random events, including $n = 0$ over some time interval $t \in [0, T]$. What does $P_0(0)$ mean and what value would it quite often take? Explain. Defend your plots.

Now let us get mathematical, yet keep in touch with reality. We will ask you to list some assumptions about a system which moves randomly from state to state, over time. Here are some questions to prompt your thinking.

2) Upon what might the probability of moving from state E_n to E_{n+1} in a given small interval $[t, t+h]$ depend? From state E_n to E_{n-1} in that same interval? From state E_n to $E_{n\pm k}$ for $k > 1$ in that same interval? Could these probabilities depend upon t, n, h? If so, articulate just how they might depend upon these quantities and why such dependence or lack thereof reflects your view of random.

So, we have thought a lot about this. Let us try to summarize the consensus thinking.

Postulates for Random or Stochastic Processes

We are not the first ones to consider randomness in this manner and it pays to look back in history - in all things. We examine what a master in probability and it application, William Feller, said in his classic text [3, p. 400].

> The physical processes which we have in mind are characterized by the two properties, (1) that they are homogeneous in time and (2) that future changes are independent of past changes. By this we mean that the forces and influences which determine the process remain absolutely unchanged, so that the probability of any particular event is the same for all time intervals of length t, independent of where this interval is situated and of the past history of the system.

Let us try to formalize our assumptions and descriptions into actionable mathematical expressions with four basic assumptions about a system which moves from state to state randomly.

C1) If the interval of time is small, say $[t, t+h]$ where h is very small, not too much can really happen. Perhaps we move from one state, say E_n to an adjacent state $E_{n\pm1}$ only. We only have states, E_n, for integers $n \geq 0$.

C2) If we are in state E_n what happens in a small time interval $[t, t+h]$ might depend only on the current state E_n and the length of the time interval h.

C3) If we are in state E_n the probability that we move to a non-adjacent state, $E_{n\pm k}$, $k > 1$ in a small time interval $[t, t+h]$ is very small. Again, not too much can happen in a tiny interval of time.

C4) It does not matter what time it is as far as whether or not the system changes state. That is, movements between states is independent of time.

You will notice that (C4) indicates that the actual time has no effect on the events. Should we allow time to shape events then we would be dealing with a much more complicated situation and we are comfortable (are you?) with what we have going for us already. These assumptions translate into axioms for a *stochastic process* as follows. Given a system which can be in any of the following states: $E_0, E_1, E_2, E_3, \ldots, E_n, \ldots$. A process which satisfies the following axioms is an example of a stochastic process and is said to be a *general birth and death process.*

1. The system changes only through one state to its neighbors (i.e. from E_n to E_{n+1} or E_{n-1} for each $n \geq 1$, but from E_0 to E_1 only.)
2. If at any time t the system is in state E_n, the probability that during the time interval $[t, t+h]$ the transition E_n to E_{n+1} occurs equals $\lambda_n h + o(h)$, and the probability that during the time interval $[t, t+h]$ the transition E_n to E_{n-1} occurs equals $\mu_n h + o(h)$. Here λ_n and μ_n reflect the possibility that these state transition probabilities might be a function of the state, E_n itself.

   Now, $o(h) << h$, i.e. $o(h)$ is very small when compared to h. Formally, $\lim_{h \to 0} \frac{o(h)}{h} = 0$. This reflects the fact that each of the probabilities we set up may not be exact, but any error is very small when compared to h, the interval length.
3. The probability that during the time interval $[t, t+h]$ more than one change occurs is $o(h)$.
4. The system's change is independent of t.

We have interest in $P_n(t)$, the probability a system is in state E_n at time t. Indeed, we have other interests beyond obtaining $P_n(t)$, e.g., $\sum_{n=0}^{\infty} n \cdot P_n(t)$, the *mean state* (average value or expected value of the state) of the system at time t, among others.

We seek to formulate a mathematical model for $P_n(t)$, the probability that the system is in state E_n at time t. We first obtain some information about $P_n(t+h)$, i.e. the probability that the system is in state E_n at time $t+h$. We derive this from the condition of the system at time t.

At time $t+h$ the system can be in state E_n if only if one of the following mutually exclusive events are is true:

1. at time t the system was in state E_n and in the interval $[t, t+h]$ no change occurred;
2. at time t the system was in state E_{n-1} and a transition to state E_n occurred in the time interval $[t, t+h]$;
3. at time t the system was in state E_{n+1} and a transition to state E_n occurred in the time interval $[t, t+h]$; or
4. at time t the system was in state E_k where $|k-n| > 1$, i.e. more than two steps from E_n and two or more transitions occurred in the time interval $[t, t+h]$.

Since the events describe in (1) - (4) above are mutually exclusive we can add their probabilities to get the probability that at time $t+h$ the system can be in state E_n. We obtain the equation (1) for $P_n(t+h)$ for $n > 0$ by adding the probabilities of these mutually exclusive and exhaustive events.

Incidentally, the product of two probabilities, e.g., $P_{n-1}(t) \cdot \lambda_{n-1}h$ signifies two independent events, i.e. one does not portend information about the other, so we just multiply their probabilities as in $P_{n-1}(t) \times \lambda_{n-1}h$ for, first, the probability that the system is in state E_{n-1} at time t and second, the probability that the system moves from state E_{n-1} to state E_n in the time interval $[t, t+h]$.

$$P_n(t+h) = P_n(t)(1 - \lambda_n h - \mu_n h) + P_{n-1}(t)\lambda_{n-1}h + P_{n+1}(t)\mu_{n+1}h + o(h)\,. \tag{1}$$

3) Describe each of the terms in (1) in words and defend in your own words just how this equation is produced. Do the same for (2) below.

For $n = 0$, we have

$$P_0(t+h) = P_0(t)(1 - \lambda_0 h - \mu_0 h) + P_1(t)\mu_1 h + o(h)\,. \tag{2}$$

We call these general *birth and death equations*.

We move to a set of differential equations (goodie!) by examining first difference quotients (here they come!) from (1) and (2).

$$P_n(t+h) - P_n(t) = -P_n(t)(\lambda_n h + \mu_n h) + P_{n-1}(t)\lambda_{n-1}h + P_{n+1}(t)\mu_{n+1}h + o(h)\,, \tag{3}$$

and then dividing by h and taking the limit as h goes to 0. Remember your calculus and the definition of the derivative. Now just what might that yield?

$$\begin{aligned}
[P_n(t+h) - P_n(t)]/h &= [-P_n(t)(\lambda_n h + \mu_n h) + P_{n-1}(t)\lambda_{n-1}h + P_{n+1}(t)\mu_{n+1}h + o(h)]/h\,, \\
&= -P_n(t)(\lambda_n + \mu_n) + P_{n-1}(t)\lambda_{n-1} + P_{n+1}(t)\mu_{n+1} + o(h)/h\,. \qquad (4)
\end{aligned}$$

And so we have (recall $\lim_{h\to 0} \frac{o(h)}{h} = 0$) differential equations for $P_n(t)$ for integer values $n > 0$:

$$P_n'(t) = dP_n(t)/dt = -P_n(t)(\lambda_n + \mu_n) + P_{n-1}(t)\lambda_{n-1} + P_{n+1}(t)\mu_{n+1}\,. \tag{5}$$

Using a similar process from (2) for $n = 0$ we can show that

$$P_0'(t) = dP_0(t)/dt = -P_0(t)(\lambda_0 + \mu_0) + P_1(t)\mu_1\,. \tag{6}$$

Here actually, $\mu_0 = 0$ for when the system is in state E_0 we cannot have a death.

There you have it – an infinite number of differential equations, one for each integer $n = 0, 1, 2, 3 \ldots$. Ask yourself, what are the initial conditions for these differential equations, i.e. $P_0(0)$ and $P_n(0)$, for $n > 1$.

Solutions to (5) and (6) and Applications

Solving (5) and (6) can prove to be difficult, especially as λ_n's and μ_n's can be quite complex as functions of n. So let us consider a motivational model.

Police Blotter or Roster

Suppose our system involves the number of names on the police blotter or roster for the evening shift, i.e. we wish to model the number of names of persons "checked-in" or who called in about something (dog barking!) at the police station during the course of the evening, beginning at midnight ($t = 0$). Our system will be in state E_n if there are n names on the roster and $P_n(t)$ is the probability that n names are on the roster at time t (say in minutes past midnight). You can make an argument as to whether or not such events are random - be prepared to do so, but for the moment let us assume these arrivals on the police roster occur randomly throughout the night.

Here are some questions to answer in this situation:

4) a) What do λ_n and μ_n signify in these cases and what must all μ_n be as there are no erasures or removals from the roster list as it is compiled?

   b) What would be a simple and reasonable assumption about λ_n's? Give a circumstance under which such a simplifying assumption might not apply.

   c) What is $P_0(0)$? What are $P_n(0)$ for all other $n > 0$?

   d) Write out (6) in this case with initial condition $P_0(0) = 1$. Then write out (5) for $n = 0, 1, 2, 3, 4$, etc. Again, what are each of these differential equation's initial conditions, i.e. what is $P_n(0)$ for $n = 1, 2, 3, 4$? Remember there could be an infinite number of these differential equations - one for each n, so we are not really trying to make this lots of work.

   e) So what we have in (d) is 5 first order differential equations. First, solve (6) to obtain $P_0(t)$.

   f) What do we know that can help us in (5) for $n = 1$? Use that information to get a friendlier (5) for $n = 1$ and then solve (5) for $n = 1$ to obtain $P_1(t)$.

   g) Now do the same to make friendlier (5) for $n = 2, 3, 4$ and solve for $P_2(t), P_3(t), P_4(t)$.

   h) Make a conjecture as to what you believe $P_n(t)$ is? If you need more evidence, then solve for $P_5(t)$, $P_6(t)$, $P_7(t)$, etc. NB: It is dangerous to live by conjecture without formal proof, but for now let us live dangerously!

   i) Using $\lambda = 3$ offer plots of $P_n(t)$ (for all values $n = 0, 1, 2, 3, 4$). Comment on the plots. What do you see and what does that mean in this situation?

   This stochastic process is known as the *Poisson process* and is used to test randomness in many phenomena as we shall see shortly.

5) Consider the expression

$$M = \sum_{n=0}^{\infty} n \cdot P_n(t)\,. \tag{7}$$

   M is the *mean or average state value of our system at time t*. It is a weighted average of values $n = 9, 1, 2, 3, \ldots$ with the percentage of time the system takes on that value, $P_n(t)$, for the respective values of n. We get a mean by multiplying each state value, $n = 0, 1, 2, 3, \ldots$, by its corresponding probability, $P_n(t)$, $n = 0, 1, 2, 3, 4 \ldots$ and adding these terms together.

Sum of Dice	Probability
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

Table 1. Events listed (sum of face on two dice) and each probability of that sum is computed.

Use Mathematica or some other technology and your conjectured general solution form for $P_n(t)$ in Activity (4) above to compute M. Then do the same, only this time use $P_n(1)$'s. Explain the meaning of $P_n(1)$ and M in this case when $t = 1$. Why is this a good number to know? How can we validate whether our model is a good one for predicting the roster size throughout the night by using M in the case with $P_n(1)$? What could we know now about these numbers $\lambda_n = \lambda$ for our roster modeling?

Incidentally, with regard to the concept of mean or average consider two dice and the event E_n where n is the sum of the faces of two six-sided, fair dice which are rolled. If the dice are fair, unrigged, then each number on each die has a $1/6$ probability of turning up and for each sum value n we give the probability that the sum of the two dice thrown would be n in Table 1. Can you figure out how each entry in the second column is computed? Hint: How can we get a sum of 5 in tossing two different dice?

We return to our police blotter modeling. Finally, consider the function (8):

$$F(t) = 1 - P_0(t) = 1 - e^{-\lambda t}\,. \tag{8}$$

$F(t)$ is the probability that "no" no arrivals take place or that at least one arrival takes place in the time interval $[0, t]$. This is often referred to as the *exponential cumulative distribution function* and is used in many areas to model first arrival (in our case first name on the police blotter or roster) or in the case of equipment, first failure.

Special (but important) case for differential equations (5), (6), and for the sum (7)

6) Suppose $\lambda_m = \lambda$ and $\mu_n = 0$ in the general birth and death process. This means there are only births (and they do not depend upon the state of the system) and there are no deaths. The system only stays at the current state or the value of the state n increases.

   Consider the expression
$$\mathcal{E} = \sum_{n=0}^{\infty} n \cdot P_n(1)\,. \tag{9}$$

   Translate this into English (and you do not need Google Translate!) You should get something like this: "The average number of births or arrivals per unit time, i.e. in the time interval $[0, 1]$." Since, one of our assumptions and something we built into our model is that the births and deaths are independent of the time, t, at which we are functioning then $\mathcal{E}$ is the average number of births in ANY and ALL time intervals of length 1. Look closely at $\mathcal{E}$ and then explain exactly what this mysterious constant λ really means.

7) Suppose now we move to a slightly different modeling scenario, $\lambda_n = n \cdot \lambda$ for $n = 0, 1, 2, 3, 4, \ldots$ and $\mu_n = 0$ for $n = 0, 1, 2, 3, , \ldots,$ and our general solutions to (5), (6) are as we found them in (4) (h) above. This will be called a *pure birth stochastic process*. Indeed, $\lambda_n = \lambda = n \cdot \lambda$ suggests that if our state E_n were the number of organisms, then the more we have the more we get or breed. We shall see how this plays out below, and $\mu_n = \mu = 0$ suggests that there is no death - at least in some short run.

   a) What do λ_n and μ_n signify in these cases?
   b) What would be a simple and reasonable assumption about λ_n's? Give a circumstance under which such a simplifying assumption might not apply.
   c) What is $P_0(0)$? What are $P_n(0)$ for all other $n > 0$?
   d) Write out (6) in this case with initial condition $P_0(0) = 0$. What then must $P_0(t)$ be? Assume $P_1(0) = 1$. This says biologically we have one organism to start our process, perhaps asexually reproducing. Then write out (5) for $n = 1, 2, 3, 4$ yup all of them. Again, what are each of these differential equation's initial conditions $P_n(0) =?$ for $n = 2, 3, 4$?
   e) So what we have in (d) is 4 first order differential equations. First, solve (5) for $n = 1$ to obtain $P_1(t)$.
   f) What do we know that can help us in (5) for $n = 2$? Use that information to get a friendlier (5) for $n = 2$ and then solve (5) for $n = 2$ to obtain $P_2(t)$.
   g) Now do the same to make (5) friendlier for $n = 3, 4, 5$ and solve for $P_3(t), P_4(t), P_5(t)$.
   h) Make a conjecture as to what you believe $P_n(t)$ is? If you need more evidence, then solve for $P_6(t)$, $P_7(t)$, $P_8(t)$, etc.
   i) Using $\lambda = 3$ offer plots of $P_n(t)$ (for all values $n = 1, 2, 3, 4$) for $t = 6$ (the first few minutes in an hour shift, wherein a fresh roster would begin. comment on the plots. What do you see and what does that mean in this situation?

Since $P_n(t)$ represents the probability that our system of pure births is in state n at time t consider the expression:

$$\mathcal{E}(\lambda, t) = \sum_{n=0}^{\infty} n \cdot P_n(t)\,. \tag{10}$$

Use Mathematica or other technology and your conjectured form of $P_n(t)$ to compute $\mathcal{E}(\lambda, t)$ in (10). So, if $P_n(t)$ represents the probability that our system of pure births is in state (or our population is of size) n at time t then $\mathcal{E}(\lambda, t)$ represents the average size of our population from births alone at time t. Look at the formula for $\mathcal{E}(\lambda, t)$ in (10) and suggest where you have seen it before.

Hint: $y'(t) = \lambda y(t)$.

If you came to the conclusion by yourself or you can read the fine print, recall the differential equation in the fine print, and solve the equation, then you have it. $\mathcal{E}(\lambda, t)$ is the same as our population model with exponential growth and growth rate of λ units per unit time per individual in the population then you are correct.

> We have a match between a probabilistic model for growth (birth only) and a deterministic model for growth, exponential growth.

Let that sink in now. This adventure down probabilistic lane and of thinking about randomness supports our deterministic assumptions early on in our study and supports the exponential growth model from a stochastic perspective.

Poisson Process Applications

We consider the results from Activity 4 and see if we can find applications. What we found (11) in Activity (4) was that for $\lambda_n = \lambda$ for $n = 0, 1, 2, 3, \ldots$ and $\mu_n = 0$ for $n = 1, 2, 3, \ldots$, then

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}\,. \tag{11}$$

In particular,

$$p_n = P_n(1) = \frac{(\lambda)^n e^{-\lambda}}{n!}\,. \tag{12}$$

We shall be paying attention to this special case, $P_n(1)$, so we give it a shorter name, p_n for the probability that we are in state E_n at time $t = 1$ which, in this case of arrivals, means that we have exactly n arrivals in one unit of time. Let us consider some situations in which this view would be appropriate.

Traffic Model

Consider the data recorded on the number of vehicles which pass a certain spot on a highway. We record the number of one minute intervals in which k vehicles pass our spot.

The question is, "Does traffic randomly pass our check point, i.e. adheres to the assumptions we are making in our stochastic process modeling?"

# vehicles (k) per minute	0	1	2	3	4	5	6	7	8	9
# minute intervals with k vehicles	2	8	15	21	19	15	10	7	1	2

Table 2. Summary of observed data on the number of one minute time intervals during which k cars pass a spot on the highway where $k = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$.

There were $0(2)+1(8)+2(15)+3(21)+4(19)+5(15)+6(10)+7(7)+8(1)+9(2) = 387$ vehicles altogether. Thus in the 100 one minute intervals we average $387/100 = 3.87$ vehicles per minute.

In attempting to model this phenomenon of traffic flow past our observation point, we ask the question: Is each interval of time equally likely to have a vehicle pass?

This may be a subtle question, for consider an observation post some 100 meters down a highway from a stop light? Do you think that in each minute of our 100 minute observation period it is equally likely that a vehicle will pass this post?

Or consider a street outside a meat packing plant in which we are observing from 4:00 PM through 5:40 PM around change of shift time? Hmmmmmmm!

Nevertheless, let us assume we are on a stretch of highway far outside of town at 9:00 AM and we take our data through 10:40 AM – 100 minutes worth.

We wish to see how well the Poisson process model fairs in this situation. This means we need to determine how well the model will predict the data we observed. We examine the Poisson model. Since the average number of vehicles per minute is $\lambda = 3.87$ we have a possible Poisson Process with

$$p_n = \frac{3.87^n e^{-3.87n}}{n!} \tag{13}$$

We evaluate these and summarize them in Table 3 for a reasonable number of values of k.

# vehicles (k) per minute	% of minute with k vehicles passing
0	0.0208584
1	0.0807219
2	0.1561970
3	0.2014940
4	0.1949450
5	0.1508880
6	0.0973226
7	0.0538055
8	0.0260284
9	0.0111922

Table 3. Summary of percentage of the one minute time intervals during which k cars pass a spot on the highway where $k = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. The percentages are the probabilities found in using the Poisson model with $\lambda = 3.87$ cars per minute.

Recall the meaning of each of these numbers. $p_2 = 0.156197$ is the percentage of the one minute time intervals which we would expect to observe exactly 2 vehicles. In our observations of 100 minutes, in how many of these minute intervals should we expect to see 2 vehicles? We should expect to see $100p_2 = 100(0.156197) = 15.6197$ or between 15 and 16 one minute time intervals in which we saw 2 vehicles pass.

Let us now compare the observed data with our model prediction to see if the Poisson process model predicts well our traffic data. We do this in Table 4. If the comparison is good, i.e. the data seem to be very close, then we can reasonably say that the number of cars passing our spot is random.

# vehicles (k) per minute	0	1	2	3	4	5	6	7	8	9
Observed # minute intervals with k vehicles	2	8	15	21	19	15	10	7	1	2
Modeled # minute intervals with k vehicles	2.1	8.1	15.6	20.1	19.5	15.1	9.7	5.4	2.6	1.1

Table 4. Summary of recorded data as to the number of one minute time intervals during which k cars pass a spot on the highway where $k = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$.

How good was our model in predicting the observed data? We predicted (did you also get the same?) somewhere between 15 and 16 intervals in which 2 vehicles were observed.

Let us take all the data from Table 4 and plot it to see how closely our model predicts the data. If they look close, then we shall say the Poisson model is a good predictor of traffic observation and traffic passing our point on the highway is random. If they do not look close, we shall say the Poisson model is a lousy predictor of traffic observation and traffic passing our point on the highway is not random.

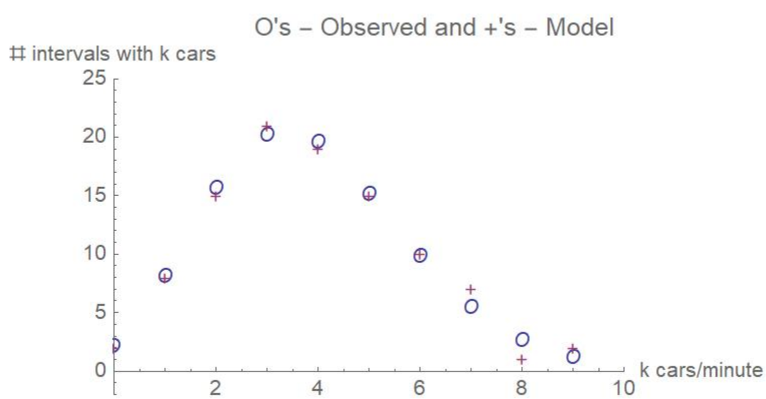

Figure 1. Plot of observed data (O's) and model data (+'s) for our Poisson process model of traffic data passing a point on the highway.

Figure 1 looks pretty good, for the two data sets are practically on top of each other. Thus we shall accept the conjecture that traffic passing a spot on the highway is random. Actually, statisticians are a bit more skeptical. For they set up statistical hypothesis, in this case, "The traffic is random." and they would only go so far as to say in this case, "'We fail to reject this hypothesis." But, that is another course!

8) Couch Potato Data

We consider Dan CouchPotato who sits in front of the TV. His only exercise is to switch channels. We have been observing him for some time and we have found the following data for the number of intervals in which he exercises k channel switches.

It is now your turn to

i) determine the average number of channel switches per minute - call this λ.
ii) Determine a theoretical Poisson p_k model for predicting the number of one minute intervals in which there are exactly k channel switches.
iii) Compare the theoretical and observed data. Does it appear that Dan CouchPotato is randomly switching channels?
iv) Discuss whether or not (and why) this Poisson model is a good model for predicting the observations.

Hint: Remember that underlying assumptions for a random process we made earlier. Go to it!

# switches (k) per minute	0	1	2	3	4	5	6	7
# one minute intervals with k switches	13	12	15	42	54	32	8	4

Table 5. Number of one minute intervals in which Dan CouchPotato made k channel switches on his TV, for $k = 0, 1, 2, 3, 4, 5, 6, 7$.

9) V-2 Rocket hits on London in World War II - A Real Application!

During World War II, London was assaulted with German flying-bombs on V-2 rockets. The British were interested in whether or not the Germans could actually target their bomb hits or were limited to random hits with their flying-bombs. In [1] the analysis which led the British to determine whether or not the Germans could target their bombs or were merely limited to random hits is presented. R. D. Clare, the author, said,

> During the flying bomb attack on London, frequent assertions were made that the points of impact of the bombs tended to be grouped in clusters. It was accordingly decided to apply a statistical test to discover whether any support could be found for this allegation.
>
> ⋮
>
> The occurrence of clustering would have been reflected ... by an excess of squares containing either a high number of flying bombs or none at all, with a deficiency in the intermediate classed. [1, p. 481]

Before we turn the analysis over to YOU, it should be noted that this analysis is very important. For if the Germans could only randomly hit targets, then deployment throughout the countryside of various security installations would serve quite well to protect them, as random bombing over a wide range was unlikely to hit a given target. However, if the Germans could actually target their flying-bombs, then the British were faced with a more potent opponent and deployment of security installations would do little to protect them.

The British mapped off the central 24 km by 24 km region of London into 1/2 km by 1/2 km square areas. Then they recorded the number of bomb hits, noting their location, and this data is in Table 6:

# bomb hits (k) per area	0	1	2	3	4	5 and over
# areas with k bomb hits	229	211	93	35	7	1

Table 6. A tally of the number of "flying-bomb" attacks on London during World War II.

Perform an analysis to determine if the bombing was random or was capable of targeting. State your assumptions and offer up a complete analysis.

10) No-hitters in Baseball - random events?

Consider the phenomenon of no-hitters in baseball. A no-hitter for a pitcher is a 9 inning game in which the pitcher allows no hits! They are rare, but are they randomly distributed? For example, do they occur more frequently near the end of the season when pitchers are at peak form or less frequently at the end of the season when pitchers are weak from throwing all season and batters or "on to their pitches"?

Consider the number of no-hitters (k) per season, and the number of seasons with k no-hitters. Using the Poisson analysis offered above, ascertain if no-hitters are randomly distributed among baseball seasons. Offer your analysis and defend your conclusions.

The following data [8, pp. 154-155] in Table 7 relates to the number of no-hitters per season for Major League Baseball from the years 1876-1989, some 114 years of professional baseball history.

# no-hitters (k) per season	0	1	2	3	4	5	6	7+ and over
# seasons with k no-hitters	26	31	23	19	10	4	1	1

Table 7. The number of seasons with k no-hitters in Major League Baseball from 1876-1989.

Question: Are no-hitters per season a random event? Offer an analysis and defend your decision on the question.

Offer a rationale for why you think your conclusion might hold, i.e. if no-hitters are random why? If not, why not?

11) Radioactive disintegration

Lord Ernest Rutherford, the famous British physicist who worked in the early part of the twentieth century, was detecting radioactive disintegrations in his laboratory. His results are reported in his book[7, p. 172] and later analyzed in [2, p. 43].

Basically Rutherford took $N = 2608$ time intervals of 7.5 seconds each and counted the number of particles in each interval which reached a counter. His data is presented in Table 8.

# particles (k) per interval	0	1	2	3	4	5	6	7	8	9	10
# intervals with k particles	57	203	383	525	532	408	273	139	454	27	16

Table 8. The number of time intervals with k particles counted in these intervals from data compiled by the physicist Ernest Rutherford.[7]

From this data, can you infer that radioactive disintegration is a random process? Write up your opinion. Defend your conclusion using the Poisson model approach.

Additional references

In the references below we offer several additional items [4, 6, 9] which the reader may find useful in developing personal confidence and competence in this area if not already present as well as seeing derivations and applications of stochastic processes.

REFERENCES

[1] Clarke, R. D. 1946. An Application of the Poisson Distribution. *Journal of the Institute of Actuaries*. 72: 481.

[2] Cramér, H. 1945. *Mathematical Methods of Statistics*. Princeton NJ: Princeton University Press.

[3] Feller, W. 1957. *Introduction to Probability Theory: Volume I, Second Edition.* New York: John Wiley and Sons. Chapter XVII.5.

[4] Hillier, S. H. and G. J. Lieberman. 2005. *Introduction to Operations Research, Eighth Edition*. New York: McGraw-Hill.

[5] Rash, A. and B. Winkel 2009. Birth and Death Process Modeling Leads to the Poisson Distribution: A Journey Worth Taking. *PRIMUS*. 19(1): 57-73.

[6] Ross, S. 2009. *Introduction to Probability Models, Tenth Edition*. Philadelphia PA: Elsevier.

[7] Rutherford, E., J. Chadwick, and C. D. Ellis. 1920. *Radiation from Radioactive Substances*. Cambridge England: Cambridge University Press.

[8] The Sporting News. 1990. Baseball no hitter data. *The Sporting News Complete Baseball Record Book*. The Sporting News: St. Louis MO.

[9] Winston, W. L. 2003. *Operations Research: Applications and Algorithms, Fourth Edition*. Independence KY: Cengage Learning.