A Quick Way to See that the Poisson Distribution is the Appropriate Mathematical Formulation for a Counting Process with Constant Rate and Intensity

Jake Bowers

November 10, 2008

There are several ways to derive the Poisson distribution. For example, it can be shown that the Poisson is a limiting case of the Binomial distribution (when the number of trials is very large and the probability of success is very small)(Ramanathan, 1993, page 62). In the case of political participation, the question is whether the Poisson is the appropriate distribution for a sequence of independent events occurring over time with a constant rate. Siegrist (2001) shows (over about 8 pages) that, if we assume that the times between the occurrence of events (the "inter-arrival times") have an exponential density function then the arrival times (the times on which events occur) have a gamma density function, and thus the sequence has a Poisson distribution.

The more direct (and shorter) proof is provided by (Ramanathan, 1993, page 64) which is like that shown by (King, 1989, page 48-50) or (Grimmett and Stirzaker, 1992, page 229). In each of these cases, the authors just begin with a process that produces counts of events over time, and end with the Poisson probability distribution. Here I explicate here a version which relies mainly on Ramanathan (1993) but uses the other proofs to clarify. Also, this proof actually shows the "induction step" alluded to but not demonstrated by the other proofs.[1].

1 Start with Numbers of Events occurring in Time

In this case one can write that $f(n,t)$ is the probability of n events occurring in a time-interval of length t. We want to know the explicit probability distribution function for $f(n,t)$. One strategy for figuring this out involves asking about the probability of n events occurring in a time-interval slightly longer than t, say, $t + \Delta t$. We can label this probability as $f(n, t + \Delta t)$. And I'll assume that this probability is equal to $\lambda \Delta t + o(\Delta t)$, where $\lambda \Delta t$ is the probability of observing one event in the time interval of length Δt and $o(\Delta t)$ is the probability of observing more than one event in that time interval. To summarize:

[1]Thanks to David Lutzer and Carl Lutzer for help on this derivation, particularly the induction step

We want to know: $f(n,t)$

We start by assuming: $f(1, t+\Delta t) = \lambda \Delta t + o(\Delta t)$

This assumption says that the probability of observing a single event is proportionate to the size of the time-interval in which observation occurs, that it is constant over time (i.e. doesn't depend on the value of t), and that it is constant over individuals. We also made a series of assumptions about what a homogeneous counting process ought to look like earlier. I repeat these assumptions here:

1. **Non-simultaneity:** Only one event can occur at any one moment.
2. **Independence:** The probability of participating in one time interval is constant and has nothing to do with the probability of participating in other time intervals.
3. **Starting from Scratch:** No events have occurred at the beginning of the period.
4. **Identical Observation Period:** The period of observation of all units is the same length of time.

Because of the assumption of non-simultaneity, we know that $o(\Delta t) = 0$, and so we can write the probability of observing one event in the time between t and Δt as just $\lambda \Delta t$, (i.e. $f(1, t+\Delta t) = \lambda \Delta t$).[2]

A count of events, n, can occur in $t + \Delta t$ in two ways given the assumptions: (1) n successes might have occurred in the first t periods $[0, t]$, with none in the last $(t, t+\Delta t)$ period; OR (2) $n-1$ successes might have occurred in the first t periods, $[0, t]$, with one success in $[t, t+\Delta t]$. Since we know the probability of observing one event in $[t, \Delta t]$ is $\lambda \Delta t$, and since we assume the events are independent, we can write the probability of only observing n events in the first $[0, t]$ interval (i.e. the first way mentioned above) as $f(n,t) \cup f(0, t+\Delta t) = f(n,t)(1 - f(1, t+\Delta t)) = f(n,t)(1 - \lambda \Delta t)$ and the probability for the second kind as $f(n-1, t) \cup f(1, t+\Delta t) = f(n-1, t) f(1, t+\Delta t) = f(n-1, t)\lambda \Delta t$. Since the probability of success in that interval is independent of the successes (or failures) occurring in other intervals, we can write the the probability of n successes in $[t, \Delta t]$ as:

$$f(n, t+\Delta t) = f(n,t)(1 - \lambda \Delta t) + f(n-1, t)\lambda \Delta t \tag{1}$$

expanding and then dividing by Δt, and rearranging terms gives:

$$\frac{f(n, t+\Delta t) - f(n,t)}{\Delta t} = \lambda f(n-1, t) - \lambda f(n,t) \tag{2}$$

[2]The assumption of non-simultaneity can actually be weakened by assuming that the probability of observing more than one event in a single instant gets very small as the length of an instant decreases so that $o(\Delta t)$ rapidly approaches 0.

2 End with a Differential Equation

The fact that equation 2 is a "quotient of differences" immediately suggests that, if we let $\Delta t \to 0$ then we get:

$$\frac{f(n,t+\Delta t)-f(n,t)}{\Delta t} \equiv \frac{df(n,t)}{dt} = \lambda f(n-1,t) - \lambda f(n,t) \tag{3}$$

So now the problem is how to solve the differential equation in 3 for $f(n,t)$. The main question is whether or not there is only one way to write $f(n,t)$ for all of the different possible values of n. One way to proceed is to check what would occur if we plugged in the first couple of possible values for n (remembering that n means "number of events").

Before we begin substituting numbers for n, we should constrain a few values of $f(n,t)$ so that this process continues to reflect the assumptions given above. Specifically, I think that $f(0,0) = 1$ and $f(n,0) = 0$ (i.e. there is no chance of an event occurring at $t = 0$). Also, $f(n,t) = 0$ for $n < 0$ (i.e. n can only take on positive values.) Given these constraints, we can solve for $f(0,t)$ by substituting $n = 0$ in equation 3:

$$\frac{df(0,t)}{dt} = -\lambda f(0,t) \tag{4}$$

$$\frac{1}{f(0,t)}\frac{df(0,t)}{dt} = -\lambda \tag{5}$$

since we know that $\int \frac{dx}{dt} = x$ it is reasonable to integrate both sides of the equation[more??]:

$$\int \frac{1}{f(0,t)}\frac{df(0,t)}{dt} = \int -\lambda dt \tag{6}$$

$$\ln f(0,t) = -\lambda t + C_0 \tag{7}$$

but $C_0 = 0$ via the non-simultaneity assumption and since no events can occur when the process starts, and C_0 represents the initial condition of the differential equation[more??], so

$$f(0,t) = e^{-\lambda t}. \tag{8}$$

Now that we have the solution for $f(0,t)$, we can try to solve for $f(1,t)$ (i.e. the probability of observing one event in an interval of length t). So for $n = 1$:

$$\frac{df(1,t)}{dt} = \lambda f(0,t) - \lambda f(1,t) \tag{9}$$

$$\frac{df(1,t)}{dt} + \lambda f(1,t) = \lambda f(0,t) \tag{10}$$

remembering equation 8, this simplifies to

$$\frac{df(1,t)}{dt} + \lambda f(1,t) = \lambda e^{-\lambda t} \tag{11}$$

Where did this derivative on the left hand side of equation 11 come from? It looks like it might have come from the use of the product rule[3] to take the derivative of $f(1,t) \cdot e^{-\lambda t}$, except not quite. However, we can make the left hand side into exactly the result of the product rule by multiplying both sides by $e^{\lambda t}$.[4] Doing this multiplication leaves us with:

$$\frac{df(1,t)}{dt}e^{\lambda t} + \lambda f(1,t)e^{\lambda t} = \lambda e^{-\lambda t}e^{\lambda t} \tag{12}$$

$$\frac{df(1,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t}f(1,t) = \lambda \tag{13}$$

Now, the left hand side really is the result of taking the derivative of $f(1,t) \cdot e^{-\lambda t}$, so we can integrate (?take antiderivatives?) of both sides:

$$\int \frac{df(1,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t}f(1,t) = \int \lambda \tag{14}$$

We know that that the left hand side is $f(1,t) \cdot e^{\lambda t}$, so the solution to this is:

$$f(1,t)e^{\lambda t} = \lambda t + C_1 \tag{15}$$

but, since we know that $f(1,0) = 0$, we can set $C_1 = 0$. So:

$$f(1,t)e^{\lambda t} = \lambda t \tag{16}$$

and, solving for $f(1,t)$ by multiplying both sides by $e^{-\lambda t}$, we finish with:

$$f(1,t) = \lambda t e^{-\lambda t} \tag{17}$$

If we had infinite time we could continue to solve for each value of n. For example, we can solve for $f(2,t)$ by following the same procedure and substituting in the solution for $f(1,t)$, and discovering that:

$$\int \frac{df(2,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t}f(2,t) = \int \lambda^2 t \tag{18}$$

$$f(2,t)e^{\lambda t} = \frac{\lambda^2 t^2}{2} \tag{19}$$

$$f(2,t) = \frac{e^{-\lambda t}\lambda^2 t^2}{2} \tag{20}$$

[3] Remember that this the rule for taking the derivative of the product of two differentiable functions:

$$\frac{d[f(x)g(x)]}{dx} = f(x)\frac{dg(x)}{dx} + g(x)\frac{df(x)}{dx}$$

[4] In the context of solving differential equations, $e^{\lambda t}$ is known as an integrating factor because ...[more??]

And for $f(3,t)$:

$$\int \frac{df(3,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t} f(3,t) = \int \frac{\lambda^3 t^2}{2} \tag{21}$$

$$f(3,t)e^{\lambda t} = \frac{\lambda^3 t^3}{6} \tag{22}$$

$$f(3,t) = \frac{e^{-\lambda t}\lambda^3 t^3}{6} \tag{23}$$

And for $f(4,t)$:

$$\int \frac{df(4,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t} f(4,t) = \int \frac{\lambda^4 t^3}{6} \tag{24}$$

$$f(4,t)e^{\lambda t} = \frac{\lambda^4 t^4}{24} \tag{25}$$

$$f(4,t) = \frac{e^{-\lambda t}\lambda^4 t^4}{24} \tag{26}$$

The preceding solutions suggest a pattern. That for any n, following equation could produce the required results (at least, it would work for $n = 0, 1, 2, 3, 4$):

$$f(n,t) = \frac{e^{-\lambda t}\lambda^n t^n}{n!} \tag{27}$$

This equation happens to be the way to write the Poisson probability density function where the mean of the function depends on time. (i.e. a simpler formulation of the distribution is $\frac{e^{-\lambda}\lambda^n}{n!}$ where λ is the mean.)

How can we prove that this is true for all n without plugging in each number one at a time? The answer, is that if we can show that this satisfies the differential equation 3 for $n+1$, then it ought to be true for n numbers since n is contained in the set of $n+1$ numbers. This is what is known as a proof by induction.

So, if equation 27 is correct for all n and $n+1$ it ought to satisfy equation 3, written as below (just as we have written it to solve for $n = 1, 2, 3$):

$$\frac{df(n+1,t)}{dt} + \lambda f(n+1,t) = \lambda f(n,t) = \frac{e^{-\lambda t}\lambda^n t^n}{n!} \tag{28}$$

multiplying both sides of the equation by the integrating factor, $e^{\lambda t}$ we get:

$$\frac{df(n+1,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t} f(n+1,t) = \frac{\lambda^n t^n}{n!} \tag{29}$$

just as we have before, observe that now the left hand side is the derivative of the product of $e^{\lambda t} \cdot f(n+1,t)$. So we can integrate both sides:

$$\int \frac{df(n+1,t)}{dt}e^{\lambda t} + \lambda e^{\lambda t} f(n+1,t) = \int \frac{\lambda^n t^n}{n!} \tag{30}$$

$$f(n+1,t)e^{\lambda t} = \frac{\lambda^{n+1} t^{n+1}}{n!(n+1)} + C_{n+1} \tag{31}$$

we can use the fact that $f(n+1, 0) = 0$ to set $C_{n+1} = 0$. That is, there are no events occurring before the first time period, and the C term here represents this initial condition. So we can go ahead and multiply both sides by $e^{-\lambda t}$ to get:

$$f(n+1, t) = \frac{e^{-\lambda t}\lambda^{n+1}t^{n+1}}{(n+1)!} \tag{32}$$

$$f(n+1, t) = \frac{e^{-\lambda t}(\lambda t)^{n+1}}{(n+1)!} \tag{33}$$

So, if equation 27 is true, you can just plug in $n + 1$ and can get the same formula (and therefore, this relation ought to work for all possible values of $n > 0$). In the end, then, we have:

$$f(n, t) = \frac{e^{-\lambda t}(\lambda t)^{n}}{n!}. \tag{34}$$

This is a Poisson process with the average rate of occurrence of events λ set to vary linearly with time t. And we got here only via making assumptions about counting events over time. $\square$

References

Grimmett, G.R. and D.R. Stirzaker. 1992. *Probability and Random Process.* 2nd ed. New York: Oxford University Press.

King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* New York: Cambridge University Press.

Ramanathan, Ramu. 1993. *Statistical Methods in Econometrics.* San Diego: Academic Press.

Siegrist, Kyle. 2001. *Virtual Laboratories in Probability and Statistics.* University of Alabama in Huntsville.