

Derivation of Poisson distribution from probability theory as logic

Andrew Fowlie

Nanjing Normal University

February 2019

1 What follows is an excerpt from ref. [1].

The Poisson distribution is usually derived as a limiting “low counting rate” approximation to the binomial distribution, but it is instructive to derive it by using probability theory as logic, directly from the statement of independence of different time intervals, using only the primitive product and sum rules. Thus define the prior information:

$$I \equiv \text{There is a positive real number } \lambda \text{ such that, given } \lambda, \text{ the probability that an event } A, \text{ or count, will occur in the time interval } (t, t + dt) \text{ is } P(A | \lambda I) = \lambda dt. \text{ Furthermore, knowledge of } \lambda \text{ makes any information } Q \text{ about the occurrence or nonoccurrence of the event in any other time interval irrelevant to this probability: } P(A | Q\lambda I) = P(A | \lambda I). \quad (1)$$

In orthodox statistics one would not want to say it this way, but instead would claim that λ is the sole causative agent present; the occurrence of the event in any other time interval exerts no physical influence on what happens in the interval dt . Our statement is very different. Denote by $h(t)$ the probability there is no count in the time interval $(0, t)$. Now the proposition:

$$R \equiv \text{No count in } (0, t + dt) \quad (2)$$

is the conjunction of the two propositions:

$$R = \text{No count in } (0, t) \cdot \text{No count in } (t, t + dt) \quad (3)$$

and so, by the independence of different time intervals, the product rule gives:

$$h(t + dt) = h(t) \cdot [1 - \lambda dt] \quad (4)$$

or $\partial h / \partial t + \lambda h(t) = 0$. The solution, with the evident initial condition $h(0) = 1$, is

$$h(t) = e^{-\lambda t}. \quad (5)$$

Now consider the probability, given A and I , of the proposition

$B \equiv$ In the interval $(0, t)$ there are exactly n counts, which happen at the times (t_1, \dots, t_n) within tolerances (dt_1, \dots, dt_n) , where $0 < t_1 < \dots < t_n < t$.

This is the conjunction of $(2n + 1)$ propositions:

$$\begin{aligned} B = & \text{No count in } (0, t_1) \cdot \text{Count in } dt_1 \cdot \\ & \text{No count in } (t_1, t_2) \cdot \text{Count in } dt_2 \cdot \\ & \dots \\ & \text{No count in } (t_{n-1}, t_n) \cdot \text{Count in } dt_n \cdot \\ & \text{No count in } (t_n, t) \end{aligned} \quad (6)$$

so by the product rule and the independence of different time intervals, the probability of this is the product of all their separate probabilities:

$$\begin{aligned} P(B | \lambda t I) = & e^{-\lambda t_1} \cdot \lambda dt_1 \cdot \\ & e^{-\lambda(t_2 - t_1)} \cdot \lambda dt_2 \cdot \\ & \dots \\ & e^{-\lambda(t_n - t_{n-1})} \cdot \lambda dt_n \cdot \\ & e^{-\lambda(t - t_n)} \end{aligned} \quad (7)$$

or, writing the proposition B now more explicitly as $B = dt_1 \dots dt_n$,

$$P(dt_1 \dots dt_n | \lambda t I) = e^{-\lambda t} \lambda^n dt_1 \dots dt_n, \quad (0 < t_1 < \dots < t_n < t) \quad (8)$$

Then what is the probability, given λ , that in the interval $(0, t)$ there are exactly n counts, whatever the times? Since different choices of the count times represent mutually exclusive propositions, the continuous form of the sum rule applies:

$$P(n | \lambda t I) = \int_0^t dt_n \dots \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 e^{-\lambda t} \lambda^n \quad (9)$$

or,

$$P(n | \lambda t I) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad (10)$$

the usual Poisson distribution. Without the time ordering in our definition of B , different choices of count times would not all be mutually exclusive events, so the sum rule would not apply in the above way.

As noted, conventional theory obtains this same formula from the premise that events in disjoint time intervals exert no physical influences on each other; the only causative agent operating is λ . Some authors have turned this around, and supposed that if we verify Eq. 10 in the frequency sense, that proves that the events were indeed causally independent!

This is an astonishing conclusion, when we note that one could design a hundred different mechanisms (or write a hundred different computer programs), which in various ways that are completely deterministic, generate the seemingly “random” data. That is, the time of the next event is completely determined by the times of the previous events by some complicated rule. Yet all of them could constrain the long-run frequencies to agree with Eq. 10 without showing any signs of correlations.

References

- [1] E. T. Jaynes. *Probability Theory as Logic*. Maximum-Entropy and Bayesian Methods. Kluwer, Dordrecht, 1990.