Poisson Arrivals See Time Averages

RONALD W. WOLFF
University of California, Berkeley, California
(Received April 1981; accepted November 1981)

In many stochastic models, particularly in queueing theory, Poisson arrivals both observe (see) a stochastic process and interact with it. In particular cases and/or under restrictive assumptions it has been shown that the fraction of arrivals that see the process in some state is equal to the fraction of time the process is in that state. In this paper, we present a proof of this result under one basic assumption: the process being observed cannot anticipate the future jumps of the Poisson process

POISSON PROCESSES are important components of many stochastic models, e.g., we often assume that customers arriving at a queue are Poisson. We can also represent exponential service in terms of a Poisson process, where Poisson "events" generate service completions only when the server is busy. For ease of exposition, we refer below to Poisson *arrivals*, usually at a queue. Our results, however, apply both to Poisson processes in other roles (e.g., departures) and to arbitrary (non-queueing) model contexts.

It is well known that in the case of the $M/G/1$ queue the fraction of arrivals who find n customers in system is equal to the fraction of time there are n customers there. More generally, a continuous-time stochastic process may be in some state (not necessarily in the Markov sense) a certain fraction of time. Suppose a Poisson stream of arrivals both observe and interact with this stochastic process, where each arrival may take the process into or out of the state. Then it will generally be true (under conditions to be specified) that the fraction of arrivals who find (observe or *see*) the process in a state is equal to the corresponding fraction of time. By *Poisson arrivals see time averages*, or simply *PASTA*, we mean the equality of these fractions.

Owing to PASTA, we can choose to solve some problems in either discrete or continuous time, e.g., for Erlang's loss formula, see Sevastyanov [1957] and Takács [1969]. What is more important, PASTA is

Subject classification: 597 Poisson process, 692 equality of time and arrival averages.

Operations Research
Vol. 30, No 2, March–April 1982

0030-364X/82/3002-0223 $01.25

crucial information in the analysis of many problems, e.g., in Avi-Itzhak et al. [1965], Melamed [1978], Nozaki and Ross [1978], Wolff [1970], and Wolff and Wrightson [1976].

To use PASTA as an essential part of an analysis, one must be able to prove it in a way that does not involve independent derivations of time and arrival averages. Proofs of this nature are in Stidham [1972], Strauch [1970], and Wolff.

Strauch's result is instantaneous; it applies to an arbitrary but fixed point on the time axis. It is not evident how to extend this result to obtain limit theorems. Wolff derived a limit theorem, but the process being observed was assumed to be stationary, and particular properties of the interaction of the Poisson stream with this process were used in the derivation. Stidham (see also Heyman and Sobel [1982]) obtains a limit theorem when the process being observed is regenerative, and also under what turn out to be unnecessary restrictions on the relation between Poisson arrivals and regeneration points. Some related results for stationary processes may also be found in Franken et al. [1981].

In this paper, we present a simple proof of PASTA. Aside from technical conditions, our only assumption, formally defined below, is that the process being observed cannot anticipate the future jumps of the Poisson process.

The proof depends on a martingale characterization of the Poisson process due to Watanabe [1964]. Hence, Equation 3 and Lemma 1 are not new. Proofs are included for completeness. Lemma 2, which is critical in applications, is new.

1. PROBLEM FORMULATION AND STATEMENT OF THE MAIN RESULT

Let $N \equiv \{N(t),\, t \geq 0\}$ be a stochastic process and $A \equiv \{A(t),\, t \geq 0\}$ be a Poisson process at rate $\lambda > 0$, both defined on some probability space $(\Omega, \mathscr{F}, P)$. $N(t)$ will represent the status of a *system* at epoch $t \geq 0$, and A an arrival process of *customers* to the system. For example, if $N(t)$ is the number of customers in system at epoch t, it will increase by one at customer arrival epochs. In general, we let $N(t)$ take on values in an arbitrary measurable space, and the interaction between A and N is unspecified.

Our objective is to compare the fraction of time N spends in some set, with the corresponding fraction of customers who see (find on arrival) the system in that set. For an arbitrary set B in the value space of N, called the *state* of N, such that $\{N(t) \in B\}$ is measurable for every $t \geq 0$, define

$$U(t) = 1 \quad \text{if} \quad N(t) \in B,$$
$$= 0 \quad \text{otherwise},$$
$$V(t) = t^{-1} \int_0^t U(s)ds,$$
$$Y(t) = \int_0^t U(s)dA(s), \qquad \text{and}$$
$$Z(t) = Y(t)/A(t).$$

We assume the sample paths of U are left continuous and have right-hand limits, w.p. 1, and that U is a jointly measurable function of $(\omega, t) \in \Omega \times [0, \infty)$.

With these definitions, $V(t)$ is the fraction of time during $[0, t]$ that N is in state B, $Y(t)$ is the number of arrivals in $[0, t]$ who find N in B, and $Z(t)$ is the fraction of arrivals in $[0, t]$ who find N in B.

In the queueing example in the first paragraph of this section, left continuity means that the arrival is not counted as being in the system at the arrival epoch. While not customary for continuous time processes, the convention of left continuity has no effect on the length of time $N(t)$ spends in any state.

The latter assumption ensures that $V(t)$ is a random variable for each t. Aside from the convention of left continuity, the sample paths of U belong to $D[0, \infty)$ w.p. 1, and have essentially the same properties. In particular, w.p. 1, U has only a finite number of discontinuities on any finite interval. Hence, for each ω in a set that has probability one, $Y(t)$ can be approximated arbitrarily closely by a function of the form

$$Y_n(t) = \textstyle\sum_{k=0}^{n-1} U(kt/n)[A((k+1)t/n) - A(kt/n)], \tag{1}$$

for sufficiently large n.

We expect the arrival process to affect the system in some way, i.e., the processes A and N are typically dependent. Consequently we expect A and U to be dependent. However, we assume the system has no *anticipation*, i.e., we do not want the future increments of A to depend on the past of U. Formally, we make the

Lack of Anticipation Assumption (LAA). For each $t \geq 0$, $\{A(t+u) - A(t), u \geq 0\}$ and $\{U(s), 0 \leq s \leq t\}$ are independent.

Our main result is

THEOREM 1. *Under LAA, $V(t) \to V(\infty)$ w.p. 1 if and only if $Z(t) \to V(\infty)$ w.p. 1, as $t \to \infty$.*

Remark 1. PASTA is true whenever convergence occurs. The only reason for making stronger assumptions, e.g., that the system is regenerative, stationary, or ergodic, is to prove convergence.

Remark 2. While $V(\infty)$ will typically be a constant, it need not be. For example, suppose N is a discrete-state Markov process with several absorbing states and initial transient state $N(0) = i$, such that i can reach each absorbing state and absorption is certain. If B is an absorbing state, $V(\infty)$ is a random variable with $P(V(\infty) \in \{0, 1\}) = 1$, where $P(V(\infty) = 1)$ is the probability of absorption in B.

Remark 3. Formally, we could have started with the process U without mentioning N. However, this would obscure the importance of Theorem 1 in applications.

2. PROOF OF THEOREM 1

From LAA, the expected value of (1) is

$$E\{Y_n(t)\} = \lambda t E\{\textstyle\sum_{k=0}^{n-1} U(kt/n)/n\}. \tag{2}$$

Now let $n \to \infty$ and apply the bounded convergence theorem.

$$E\{Y(t)\} = \lim_{n\to\infty} E\{Y_n(t)\} = \lambda t E\{V(t)\} = \lambda E\left\{\int_0^t U(s)ds\right\}, \tag{3}$$

i.e., on any finite interval, the expected number of arrivals who find the system in state B is equal to the arrival rate times the expected length of time it is there. Not only is this result crucial for what follows, it is interesting in its own right.

Remark 4. Conditioned on $A(t)$, the conditional expectations corresponding to (3) are typically not equal. For example, let $A(t)$ be the arrival process and $N(t)$ the number of customers in system at epoch t for an $M/M/1$ queue, with $N(0) = 0$. Let B denote that the system is busy. Then $E\{Y(t)|A(t) = 1\} = 0$, but $E\{V(t)|A(t) = 1\} > 0$.

Now define the process R,

$$R(t) = Y(t) - \lambda t V(t), \qquad t \geq 0. \tag{4}$$

Theorem 1 is an easy consequence of the following two lemmas.

LEMMA 1. *R is a continuous-time martingale.*

Proof. We need to show that for $t \geq 0$, $h > 0$, the increments of R have the property

$$E\{R(t+h) - R(t)|R(s), \qquad 0 \leq s \leq t\} = 0. \tag{5}$$

Let $\mathscr{F}_t$ be the σ-field generated by $\{A(s), U(s); 0 \leq s \leq t\}$. From LAA and the independent-increments property of the Poisson process, we may repeat the argument used to derive (3) for any set in $\mathscr{F}_t$:

$$E\{Y(t+h) - Y(t)|\mathscr{F}_t\} = \lambda E\left\{\int_t^{t+h} U(s)ds|\mathscr{F}_t\right\}. \qquad (6)$$

Because the σ-field generated by $\{R(s), 0 \le s \le t\}$ is a subset of $\mathscr{F}_t$, (5) is immediate.

Remark 5. If we assume LAA, that R is a martingale, and that A is an increasing right-continuous function with jumps of size one—Watanabe has shown that A is a Poisson process. Hence the martingale property is a characterization of a Poisson process.

LEMMA 2. *$R(t)/t \to 0$ w.p. 1 as $t \to \infty$.*

Proof. From $Y(t) \le A(t)$ and $V(t) \le 1$, observe that

$$E\{R^2(t)\} \le E\{A^2(t)\} + \lambda^2 t^2 = \lambda t + 2\lambda^2 t^2 \equiv k(t). \qquad (7)$$

From Lemma 1, the process $\{X_n\}$ defined by

$$X_n = R(nh) - R((n-1)h), \qquad n = 1, 2, \cdots,$$

is a discrete-time martingale for any $h > 0$.

By the same bounding argument used to obtain (7),

$$E(X_n^2) \le k(h), \qquad \text{and} \qquad (8)$$

$$\textstyle\sum_{n=1}^{\infty} E(X_n^2)/n^2 < \infty. \qquad (9)$$

From (9),

$$R(nh)/n \to 0 \text{ w.p. } 1, \qquad (10)$$

by Theorem 3, p. 243, of Feller [1971]. Observe that because $Y(t)$ is monotone and changes in $tV(t)$ on an interval of length h are bounded by h,

$$R(nh) - \lambda h \le R(t) \le R((n+1)h) + \lambda h, \qquad t \in [nh, (n+1)h]. \quad (11)$$

Now divide (11) by t and let $t \to \infty$, where $n(t) \equiv [t/h] \to \infty$. The Lemma follows from (10).

Finally, Theorem 1 follows immediately from Lemma 2 and $A(t)/t \to \lambda$ w.p. 1 by writing

$$R(t)/t = (Y(t)/A(t))(A(t)/t) - \lambda V(t).$$

Remark 6. Under LAA, Lemma 2 implies other versions of Theorem 1 under weaker modes of convergence, e.g., in probability or in distribution. An expected-value version follows directly from (3).

3. NONSTATIONARY POISSON ARRIVALS

When the rate of the Poisson process $A(t)$, $\lambda(t)$, is a function of time, Theorem 1 may easily be generalized. (For a generalization of the martingale characterization of Watanabe, see Bremaud and Jacod [1977].) Assume $\lambda(t)$ is bounded and integrable, and that the following limit exists:

$$\lim_{t\to\infty} t^{-1}\Lambda(t) \equiv \bar{\lambda}, \qquad 0 < \bar{\lambda} < \infty,$$

where $\Lambda(t) = \int_0^t \lambda(s)ds$, $t \geq 0$. Define an "arrival rate weighted time average":

$$\bar{V}(t) = \Lambda(t)^{-1} \int_0^t U(s)\lambda(s)ds, \qquad t \geq 0.$$

Otherwise, all assumptions and the notation are the same as before.

Analogous to (3), we now have

$$E\{Y(t)\} = \int_0^t E\{U(s)\}\lambda(s)ds, \tag{12}$$

and, by essentially the same argument.

THEOREM 2. *Under LAA,* $\bar{V}(t) \to V(\infty)$ *w.p.* 1 *if and only if* $Z(t) \to V(\infty)$ *w.p.* 1, *as* $t \to \infty$.

Remark 7. The $M/G/1$ queue with nonstationary Poisson arrivals, where $\lambda(t)$ is a periodic function, was studied by Harrison and Lemoine [1977]. Their relation between distributions G and H on p. 567 follows from Theorem 2.

4. CONCLUSION

It is the independent-increments property of the Poisson process that really accounts for PASTA. For either stationary counting processes or renewal processes, only the (possibly compound) Poisson process has this property. Because other arrival processes in effect "anticipate themselves," LAA will typically be false.

Nevertheless, instances do occur where non-Poisson arrivals see time averages. In particular, for the $M/M/1$ queue with feedback, Burke [1976] has shown that the composite stream of exogenous Poisson arrivals and feedback customers is not Poisson, even though it is easily shown that this stream sees time averages. Apparently, some of the simple state probability results for certain queueing networks are accounted for by non-Poisson composite arrival streams that see time averages. Interestingly, Burke uses PASTA in his proof. For his model, this property is of course well known.

We close with two examples. Example 1 illustrates the versatility of PASTA in applications. Example 2 presents a situation where, for a model of some importance, PASTA is false.

Example 1. The $GI/M/c$ queue with arrival rate λ and service rate μ/server. Let π_n be the fraction of arrivals that find n customers in system, and let p_n be the corresponding fraction of time, $n = 0, 1, \cdots$.

We want to show the following well known result:

$$\mu_n p_n = \lambda \pi_{n-1}, \qquad n = 1, 2, \cdots, \tag{13}$$

where $\mu_n = \mu \min(c, n)$. We shall show (13) by equating transition rates (number of transitions per unit time) across the boundary between states n and $n - 1$. Clearly, $\lambda \pi_{n-1}$ is the transition rate from $n - 1$ to n.

To find the transition rate in the reverse direction, let $\{A(t)\}$ be a Poisson event process at rate μ_n that generates departures when the number of customers in system is n, i.e., at each of these events, a departure occurs if and only if the number in system is n. (In effect, we define a separate Poisson process for each state.) Let $Y(t)$ be the number of transitions from n to $n - 1$ in $[0, t]$. We have $A(t)/t \to \mu_n$ w.p. 1 and, from PASTA, $Y(t)/A(t) \to p_n$ w.p. 1. Hence,

$$Y(t)/t = [Y(t)/A(t)][A(t)/t] \to \mu_n p_n \text{ w.p. } 1,$$

the transition rate from n to $n - 1$, and we have (13). Extension: For $\{\mu_n\}$ an arbitrary function n, (13) is true.

Example 2. Two single-channel queues in tandem with Poisson arrivals and exponential service, with the usual independence assumptions except that for each customer, the service times at each station are the *same* random variable. Several recent papers, e.g. Pinedo and Wolff [1982], consider this model, because it captures an important feature of data transmission systems.

It is well known that the departure process from an $M/M/1$ queue is Poisson. Thus the departure process from the first station is a Poisson arrival process at the second. We shall compare time and arrival averages at the *second* station, where $\{A(t)\}$ is the arrival process at that station and, for every $t \geq 0$, $N(t)$ is the *work* there, i.e., $N(t)$ is the sum of the service times of all customers in queue and the remaining service time of the customer in service, if any, at the second station. Brumelle [1971] has shown that the time average work, $E(N)$, has the following relation to D, a customer's delay in queue:

$$E(N) = \lambda E(DS) + \lambda E(S^2)/2, \tag{14}$$

where S is a customer service time, and D and S refer to the *same* customer. Analogous to "$L = \lambda W$," $E(N)$ is a time average and the right-hand side of (14) is an average over customers. See Brumelle for formal definitions. Because the second station is single-channel (and customers are served in their order of arrival), $E(D)$ is the average work seen by an arrival.

Because this model is regenerative, it is easily shown that $E(N)$ and

$E(D)$ are the respective means of the time average and arrival average work *distributions.* If PASTA were true, these distributions would be the same, and we would have $E(N) = E(D)$.

It is easily seen that a customer's delay at the second station is a decreasing function of that customer's service time at the first. Hence D and S are negatively correlated, and

$$E(DS) < E(D)E(S). \tag{15}$$

Suppose for a moment that PASTA is true. From PASTA, (14), and (15), we easily obtain

$$E(D) = E(N) < \lambda E(D)E(S) + \lambda E(S^2)/2,$$

$$E(D) < \lambda E(S^2)/2(1 - \lambda E(S)). \tag{16}$$

Specializing to exponential service, the expression on the right-hand side of (16) is the expected delay in queue for the standard $M/M/1$ queue. This would be the expected delay in queue at the second station if service times of the same customer were independent. However, Pinedo and Wolff show that under light traffic (small $\lambda E(S)$), the inequality in (16) is reversed. Hence PASTA is false, at least in light traffic. On reflection, and, from Theorem 1, the only alternative, we see that LAA fails to hold.

ACKNOWLEDGMENT

I wish to thank Ward Whitt for suggestions that led to substantial improvements on the original version of this paper. The most important change is the statement of LAA. He also observed that the method of proof leads, without extra effort, to an "if and only if" conclusion in Theorem 1. Finally, he suggested the generalization in Section 3. I wish also to thank the referee of the original version for similar suggestions regarding LAA, the important observation that a device used in the original version was not needed, and other constructive comments. I wish to thank Erhan Çinlar for pointing out the connection between results here and the work of Watanabe.

The original version of this paper was written while the author was visiting Bell Telephone Laboratories, Holmdel, New Jersey. This revised version was written while the author was a Visiting Scientist at the National Research Institute for Mathematical Sciences, CSIR, Pretoria, South Africa.

REFERENCES

AVI-ITZHAK, B., W. L. MAXWELL AND L. W. MILLER. 1965. Queuing with Alternating Priorities. *Opns. Res.* **13,** 306–318.

BREMAUD, P., AND J. JACOD. 1977. Processes ponctuels et martingales: résults récents sur la modélisation et le filtrage. *Adv. Appl. Prob.* **9,** 362–416.

BRUMELLE, S. L. 1971. On the Relation between Customer and Time Averages in Queues. *J. Appl. Prob.* **8,** 508–520.

BURKE, P. J. 1976. Proof of a Conjecture on the Interarrival—Time Distribution in an $M/M/1$ Queue with Feedback. *IEEE Trans. Comm.* **24,** 575–576.

FELLER, W. 1971. *An Introduction to Probability Theory and Its Applications, Vol. II,* Ed. 2. Wiley, New York.

FRANKEN, P., D. KÖNIG, U. ARNDT AND V. SCHMIDT. 1981. *Queues and Point Processes.* Akademie-Verlag, Berlin.

HARRISON, J. M., AND A. J. LEMOINE. 1977. Limit Theorems for Periodic Queues. *J. Appl. Prob.* **14,** 566–576.

HEYMAN, D. P., AND M. J. SOBEL. 1982. *Stochastic Models in Operations Research.* McGraw-Hill, New York.

MELAMED, B. 1978. On Markov Jump Processes at Traffic Epochs and Their Queueing-Theoretic Applications. Department of Industrial Engineering and Management Science, Northwestern University (to appear in *Mathematics of Operations Research*).

NOZAKI, S. A., AND S. M. ROSS. 1978. Approximations in Finite-Capacity Multi-Server Queues with Poisson Arrivals. *J. Appl. Prob.* **15,** 826–834.

PINEDO, M. L., AND R. W. WOLFF. 1982. A Comparison between Independent and Dependent Service Times in Tandem Queues (to appear in *Operations Research*).

SEVASTYANOV, B. A. 1957. An Ergodic Theorem for Markov Processes and its Application to Telephone Systems with Refusals. *Theor. Prob. Appl.* **2,** 104–112.

STIDHAM, S., JR. 1972. Regenerative Processes in the Theory of Queues, with Applications to the Alternating Priority Queue. *Adv. Appl. Prob.* **4,** 542–577.

STRAUCH, R. E. 1970. When a Queue Looks the Same to an Arriving Customer as to an Observer. *Mgmt. Sci.* **17,** 140–141.

TAKÁCS, L. 1969. On Erlang's Loss Formula. *Ann. Math. Statist.* **40,** 71–78.

WATANABE, S. 1964. On Discontinuous Additive Functionals and Levy Measures of a Markov Process. *Jpn. J. Math.* **34,** 53–70.

WOLFF, R. W. 1970. Work-Conserving Priorities. *J. Appl. Prob.* **7,** 327–337.

WOLFF, R. W., AND C. W. WRIGHTSON. 1976. An Extension of Erlang's Loss Formula. *J. Appl. Prob.* **13,** 628–632.