REVUE FRANÇAISE D'AUTOMATIQUE, INFORMATIQUE, RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

M. KRAKOWSKI

Arrival and departure processes in queues. Pollaczek-Khintchine formulas for bulk arrivals and bounded systems

Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle, tome 8, n° V1 (1974), p. 45-56

<http://www.numdam.org/item?id=RO_1974__8_1_45_0>

R.A.I.R.O.
(8e année, V-1, 1974, p. 45-56)

ARRIVAL AND DEPARTURE PROCESSES IN QUEUES. POLLACZEK-KHINTCHINE FORMULAS FOR BULK ARRIVALS AND BOUNDED SYSTEMS

par M. KRAKOWSKI

Abstract. — *This paper derives several relations among the probabilities of queue size at instances of arrival, departure, and random obsevation. This is done for $G/G/c$, $M/G/c$ without or with bulk arrivals, $G/M/c$, bounded $M/G/c$, and birth-and-death systems.*

In the Supplement the Pollaczek-Khintchine formula is extended to bounded $M/G/1$ and to $M/G/1$ with bulk arrivals.

Several equalities and inequalities relate the expected queue sizes to the sizes expected by an arrival.

In the theory and practice of queuing systems it is often necessary to distinguish among the following distribution functions.

$P(n, t)$ = probability of n customers being in the system at time t;

$p(n, t)$ = probability of n customers being in the queue at time t;

$Q(n, t)$ = probability that a new arrival at time t finds n other customers in the system; *cf. (A.3)*

$Q^*(n, t)$ = probability that a departing customer (serviced or reneging) at time t leaves behind n other customers in the system; *cf. (A.4)*

$q(n, t)$ = probability that a new arrival at time t finds n other customers in the queue;

$q^*(n, t)$ = probability that a customer leaving the queue (entering the service booth or reneging) at time t leaves behind n other customers in the queue.

Note that the term « queue » denotes the waiting line for service only; the term « queuing system » includes both the waiting customers and those being serviced.

When the time t does not enter explicitly in the distribution functions, as in the stationary case, the variable t will be omitted and the variable n will be lowered into the subscript level. Thus $P(n, t)$ becomes P_n, etc.

Notation for the stationary regime

L = expected number of customers in the queuing system $= \sum_{1}^{\infty} nP_n$;

l = expected number of customers in the queue $= \sum_{1}^{\infty} nP_{n+1}$;

Q = expected number of (other) customers found in the system by a new arrival;

Q^* = expected number of customers left behind in the system by a departing customer (serviced or reneging);

q = expected number of (other) customers found in the queue by a new arrival;

q^* = expected number of customers left behind in the queue by a customer leaving the queue (entering the booth or reneging);

c = number of channels;

λ = frequency of arrivals; if this frequency depends on the state of the system n it will be shown as λ_n;

μ = frequency of departures for a channel under full-load conditions, i.e. $1/\mu$ = average servicing time; when this frequency depends on the state of the system n it will be shown as μ_n;

$\rho = \lambda/\mu$.

In the notation $G/G/c$, $M/G/c$, $M/M/c$, etc. the first letter refers to the input, the second to the service times and the third to the number of channels. « M » denotes Markovian, that is either Poissonian input or negative-exponential times; « G » stands for general, independent interarrival or service intervals.

The following Theorems, A through E, appear to be little known, excepting Theorem B for the case or a single channel. It is the object of this communication to provide simple proofs of these statements, all referring to stationary conditions.

Equations relating L to Q and l to q are given for $G/M/c$ and $M/G/c$. In the Supplement, Section 2, an inequality between l ansd q is derived for $G/G/1$. In section 3 the Pollaczek-Khintchine formula is extended, using theorems B and D, to bounded $M/G/1$. With the help of scholion to theorem B, it is extended to $M/G/1$ with bulked arrivals in Section 4.

The proof of theorem E has an example of two processes with pairwise identical state probabilities P_n but with pairwise different encounter probabilities Q_n.

The proofs are not rigorous — we stress the perceptible and intuitive — and may be treated as plausibility arguments.

Theorem A

Under very wide conditions, in particular for $G/G/c$ where, c can be a random variable, and for the birth-and-death process we have

$$Q_n = Q^*_n, \quad \text{and hence} \quad Q = Q^*, \tag{A.1}$$

and

$$q_n = q^*_n, \quad \text{and thus} \quad q = q^*. \tag{A.2}$$

According to (A.1) the probability that a newcomer (a bona fide customer, not an outside observer) finds n other customers in the system equals the probability that a departing customer leaves behind n others in the system. The expected size Q is the size expected just prior to an arrival and just after a departure.

(A.2) makes similar statements about the queue (system minus service stations). Departures, in both (A.1) and in (A.2) can be due to completed service or to reneging of customers.

Moreover, the system may be bounded and the number of servers may be a random variable. It is, however assumed in theorem A that no bulking takes place, i. e. that the probability of the simultaneous occurence of two or more events is zero (or that the probability of two or more events during dt is of the order dt^2); an event is an arrival or a departure of a customer, possibly by reneging. If bulking does take place the theorem has to be modified but we shall not do it here. (For some purposes a strategem such as resolving a multiple event artificially into a sequence of single events may restore the applicability of theorem A in its stated form.) Note, however, the Scholion to theorem B and our second extension of the Pollaczek-Khinchine Theorem which deal with bulked arrivals.

Scholion. If there are several classes of customers, each class with its own interarrival and service distribution functions, then theorem A applies separately to each class of customers.

Theorem B

For a system $M/G/c$, where the number of channels, c, may be a random variable and where reneging may take place, we have

$$P_n = Q_n = Q^*_n \quad \text{and} \quad L = Q = Q^*. \tag{B.1}$$

Similarly

$$p_n = q_n = q^*_n \quad \text{and} \quad l = q = q^*. \tag{B.2}$$

(B.1) means that for the oft treated system $M/G/c$ the probability of finding n other customers in the system by a (bonafide) new customer equals the probability of finding n customers by an outside observer at a random instant of time; and the two probabilities equal the probability that a departing customer leaves behind n other customers. The three corresponding average system sizes are all equal.

Scholion. Consider now bulked arrivals, where the multiple arrival events form a Poisson process, and the number within the incoming group may be a random variable. Then the probability that an arriving group finds n customers within the system (queue) equals $P_n\,(p_n)$ i.e. the probability that a random observation will find n customers in the system (queue). This generalizes the equation $P_n = Q_n$ and the equation $p_n = q_n$. The relations $L = Q$ and $l = q$ also hold for the bulked Poisson input. Note that in this extension we make no statements about the size of the system at departure instants.

If there are several classes of customers then (B. 1), (B. 2) and the Scholion apply to each class separately.

Theorem C

For $G/M/c$ we have

(C.1) $$\rho Q_n = \min\,[c\,;n+1]P_{n+1}$$

(C.2) $$cL = \rho(Q+1) + \sum_{1}^{c} k(c-k)P_k.$$

When $c = 1$, i.e. for a system $G/M/1$, we have

(C.3) $$\rho Q_n = P_{n+1}$$

and

(C.4) $$L = \rho(Q+1) \quad \text{and} \quad l = \rho Q.$$

Theorem D

If the system $M/G/c$ is limited to N customers (that is in state N the arrivals drop out without repeating their attemps; or the frequency of arrivals is zero when the system is full and is λ otherwise) then

(D.1) $$Q_n = P_n/(1-P_N) \quad , \quad q_n = p_n/(1-P_N) \quad ; \quad 0 \leqslant n \leqslant N-1;$$

(D.2) $$Q = L/(1-P_N) \quad \text{and} \quad q = l/(1-P_N).$$

(D.1) States that the (successful) probabilities of encounter, Q_n, where $0 \leqslant n \leqslant N-1$, are obtained by normalizing the state probabilities P_n to add up to one. (D. 2) follows from (D. 1).

$Q_N = 0$, but $P_N \neq 0$, of course. Thus,

$$Q_m/Q_n = P_m/P_n$$

for n and m ranging from 0 to $N-1$, and

$$\sum_0^{N-1} Q_n = 1 \quad \text{and} \quad \sum_0^{N-1} P_n = 1 - P_N.$$

When N tends to infinity (D.1) and (D.2) become (B.1) and (B.2), respecively.

As in theorem B, the relation proved in theorem A, namely $Q_n = Q^*_n$ still holds, of course.

Theorem E

For the birth-and-death process let $\lambda_n \mathrm{d}t$ and $\mu_n \mathrm{d}t$ be the probabilities of an arrival and of a departure, respectively, within the time $\mathrm{d}t$, when the system is in state n; then

$$\text{(E.1)} \qquad Q_n = \lambda_n P_n/\lambda, \quad \text{where} \quad \lambda = \sum_0^{\infty} \lambda_n P_n$$

and

$$\text{(E.2)} \qquad Q_n = Q_0 \prod_1^n \frac{\lambda_i}{\mu_i}.$$

Moreover,

$$\text{(E.3)} \qquad Q_n = P_n \quad \text{if, and only if,} \quad \lambda_n = \lambda.$$

As shown by an example in the proof, two birth-and-death processes may have identical sets of state probabilities, P_n, and different sets of encounter probabilities Q_n.

Proof of Theorem A

The formal definitions of $Q(n, t)$ and of $Q^*(n, t)$ are,

$$\text{(A.3)} \quad Q(n, t) = \lim_{\mathrm{d}t \to 0} \text{Prob} \left[\begin{array}{c|c} n \text{ customers} & \text{new arrival} \\ \text{at time } t & \text{within } (t, t + \mathrm{d}t) \end{array} \right]$$

$$= \lim_{\mathrm{d}t \to 0} \frac{\text{Prob}\,[n \text{ at } t \;\&\; \text{an arrival within } (t, t + \mathrm{d}t)]}{\text{Prob}\,[\text{a new arrival within } (t, t + \mathrm{d}t)]}$$

$$\text{(A.4)} \quad Q^*(n, t) = \lim_{\mathrm{d}t \to 0} \text{Prob}\,[n + 1 \text{ at } t \mid \text{a departure within } (t, t + \mathrm{d}t)]$$

$$= \lim_{\mathrm{d}t \to 0} \frac{\text{Prob}\,[n + 1 \text{ at } t \;\&\; \text{a departure within } (t, t + \mathrm{d}t)]}{\text{Prob}\,[\text{a departure within } (t, t + \mathrm{d}t)]}$$

Under stationary regime and no bulking the numerators in (A. 3) and (A. 4) must be equal since the probability of a transition from state n to the state $n + 1$ during dt equals the probability of a transition from state $n + 1$ to the state n during dt; this is the same as saying that the frequency of transitions from n to $n + 1$ must equal to the frequency of transitions from $n + 1$ to n under stationary conditions.

The denominators in (A. 3) and in (A. 4) must also be equal under stationary conditions since the frequency of arrivals into the system must equal the frequency of departures.

(In the Scholion to theorem B (A. 3) is adapted to bulked arrivals.)

Therefore

$$Q_n = Q^*_n \quad \text{and} \quad Q = Q^* \tag{A.1}$$

as was to be shown. The proof of (A.2) is quite similar to that of (A.1) and will be omitted.

Notice that in the above considerations it was immaterial whether a departure was a result of completed service or of reneging. In view of the simplicity of the assumptions, stationarity of operation and no multiple events, it is clear why the theorem A holds under such wide conditions.

The proof of the Scholion to theorem A is virtually a rephrasing of the proof above for a single type of customers.

Alternate Proof of Theorem A

Consider an enclosure, e.g. a waiting room, with people moving in and out, singly, under stationary regime. The event « an arrival encounters n other people » and the event « a departer leaves behind n people » must alternate. Hence their frequencies Q_n and Q_n^* are equal.

This scenario includes the queuing system, or the queue only, of $G/G/c$ with random or programmed c, reneging custoners and servers, and dependent inter-event intervals. Coincidence of two events is assumed of probability zero.

The nature of the argument shows that the system need be stationary in a very weak sense. The inside of a scheduled airplane is stationary enough.

(If bulked arrivals or departures take place then the event « an arrival group increases the system size to more than n » and the event « a departure group decreases the size of the system to less than $n + 1$ » must alternate. This is clear geometrically also. A cut separating the system sizes « n or less » from the system sizes « more than n » is crossed as frequently upwards as downwards in the course of its history.)

Proof of Theorem B

For $M/G/c$ the numerator in (A.3) above is, under stationary conditions, $P_n \cdot \lambda \cdot \mathrm{d}t$; the denominator is $\lambda \cdot \mathrm{d}t$. Hence, $Q_n = P_n$, and using the result of theorem A we get $Q_n = Q^*_n = P_n$. It follows therefore that also $Q = Q^* = L$. This completes the proof of (B.1). The proof of (B.2) is quite similar and will be omitted here.

The result $Q_n = P_n$ is intuitively plausible. A poissonian arrival is « random » and can be thought of as the investigator's probing instant. The Poisson arrival process of intensity λ can be thought of as composed of two independent Poisson processes with intensities $\mathrm{d}\lambda$ and $\lambda - \mathrm{d}\lambda$. The first process $\mathrm{d}\lambda$, will be used only as the surveyor's timing process which he uses the way a statistician uses a table of random numbers. The observations are made at the instances of the events of the process $\mathrm{d}\lambda$, without any « customers » joining then the queuing system. When $\mathrm{d}\lambda$ tends to zero the observations tend to become successively independent, while $\lambda - \mathrm{d}\lambda$ tends to the original process λ, yielding in the limit the state probabilities P_n. (The same reasoning does not apply to a general arrival process because this process cannot, generally, be represented as a sum of two stochastically independent processes $\mathrm{d}\lambda$ and $\lambda - \mathrm{d}\lambda$, where $\mathrm{d}\lambda$ is poissonian.)

To prove the Scholion to theorem *B* notice that if in (A.3) the word « arrival » is replaced by » group of arrivals » the equality $P_n = Q_n$ remains valid.

The proof that theorem B applies to each class of customers separately, when there are several classes of them, is obtained by applying the above reasoning to each class separately and using the Scholion to theorem A.

Proof of Theorem C

The numerator in (A. 3) is the expected number of transitions from state n to state $n+1$ during $\mathrm{d}t$. Under stationary regime this equals the expected number of transitions from state $n+1$ to state n which for $G/M/c$ is

$$P_{n+1}\,\mu \min\,[c\,;n+1]\,\mathrm{d}t.$$

The denominator is $\lambda \mathrm{d}t$, since the overall arrival rate is equal to the overall departure rate. Therefore,

$$\rho Q_n = P_n \min\,[c\,;n+1].$$

Multiplying both sides of the above equality by $n+1$ and summing from $n = 0$ to infinity we get

$$(Q+1)\rho = cL - \sum_{1}^{c} k(c-k)P_k.$$

When $c = 1$ we have $L = (Q+1)\rho$.

Proof of Theorem D

The numerator in (A.3) is now (the system being $M/G/c$ and limited to N customers, where $N \geqslant c$), under stationary conditions,

$$P_n \lambda \mathrm{d}t \text{ when } n \leqslant N - 1, \text{ and equal to zero when } n = N.$$

The denominator in (A.3) is, counting successful arrivals only,

$$\lambda[1 - P(N)]\mathrm{d}t,$$

since only when $n = N$ are new arrivals barred from the system. Hence,

$$Q_n = P_n/[1 - P_N], \qquad \text{and} \qquad Q_m/Q_n = P_m/P_n;$$

the integers m and n range from 0 to $N - 1$ and $\sum_0^{N-1} Q_k = \sum_0^N P_k = 1$.

Proof of Theorem E

The probability of an arrival during $\mathrm{d}t$ is $\lambda_n \mathrm{d}t$ when the system is in state n, and the probability of departure is then $\mu_n \mathrm{d}t$.

The numerator in (A.3) is $P_n \lambda_n \mathrm{d}t$ and the denominator is $\sum_0^{\infty} P_n \lambda_n = \lambda$, the average overall arrival rate.

Therefore

$$Q_n = \lambda_n P_n/\lambda, \tag{E.1}$$

and

$$Q_{n+1}/Q_n = \lambda_{n+1} P_{n+1}/\lambda_n P_n.$$

Since transitions from state n to $n + 1$ are as frequent as those from state $n + 1$ to n, we have

$$P_n \lambda_n = P_{n+1} \mu_{n+1} \tag{E.2}$$

and it follows from the last two equations that

$$Q_n = Q_0 \prod_1^n \lambda_i/\mu_i. \tag{E.3}$$

In view of theorem A $Q_n = Q^*_n$.

Remarkably, it follows that two birth-and-death processes may have identical state probabilities P_n and different encounter probabilities Q_n. If process $\# 1$ has λ_n, μ_n while process $\# 2$ has $\tilde{\lambda}_n = \sigma$, $\tilde{\mu}_n = \sigma \mu_n/\lambda_{n-1}$, where σ is a constant frequency, then $\tilde{P}_n = P_n$ since $\tilde{\mu}_{n+1}/\tilde{\lambda}_n = \mu_{n+1}/\lambda_n$ in (E.2).

Furthermore $\tilde{Q}_n = P_n$ as follows from (E.1), but Q_n is different from P_n when $\lambda_n \neq \lambda$ in (E.1).

Supplement : Derivation of the Pollaczek-Khintchine Formula for Expected Queue Sizes and Extensions

Section 1

For a new arrival in a $G/G/1$ system the expected waiting time for the beginning of service is composed of :

a) the expected remaining service time of the station occupant, if the system is not empty, and

b) the expected servicing time of the queue encountered by the newcomer.

The duration *a*) is $(1 - Q_0)R$ where the factor $(1 - Q_0)$ is the probability that the arrival encounters a busy service station, and R is defined as the expected remaining service time of the booth occupant at the instant of an arrival, conditional upon the system being non-empty. The duration *b*) is q/μ, of course. Therefore, the waiting time, w, for service is

$$w = (1 - Q_0)R + q\,\frac{1}{\mu}. \tag{S.1}$$

Since, as well known (cf. ref. 1, section 2, or ref. 2)

$$l = \lambda w \tag{S.2}$$

we get from (S.1) and (S.2)

$$l = \rho q + \lambda(1 - Q_0)R. \tag{S.3}$$

Specializing the system $G/G/1$ to $M/G/1$ we have, as stated in theorem B

$$q = l \quad \text{(cf. B.2)}$$

and

$$Q_0 = P_0 \quad \text{(cf. B.1)}.$$

Furthermore, with a Poisson input, the remaining service time R becomes a characteristic of the service time distribution only. It is then the expected remaining service time at a « random » instant, or in a demographic interpretation, the average remaining lifetime (or average « age », by symmetry) in a stationary population of occupants of service stations. Therefore, recalling that for $G/G/1$ $1 - P_0 = \rho$, we get for $M/G/1$

$$l = \rho l + \lambda\rho R \tag{S.4}$$

and

$$l = \lambda\rho R/(1 - \rho). \tag{S.5}$$

Notice that for $M/M/1$ $R = \frac{1}{\mu}$ and $l = \frac{\rho^2}{1-\rho}$, as well known. (S.5) is the Pollaczek-Khinchin formula for waiting line expectation in its linear form (linear with respect to R) (cf. ref. 1).

To obtain the usual Pollaczek-Khintchine equation we have to represent R in terms of the variance and the expectation of the service time durations.

This was done in ref. 1, p. 75, equation (4.6) :

$$R = (\text{var } \tilde{X} + X^2)/2X \tag{S.6}$$

where $\tilde{X}$ = total service time (random variable) and $X = E(\tilde{X}) = 1/\mu$.

Section 2

From (S.3) it follows that for $G/G/1$

$$l \geqslant \rho q. \tag{S.7}$$

If $R \leqslant 1/\mu$ then also

$$l \leqslant \rho q + \rho(1 - Q_0) \leqslant \rho q + \rho \quad \text{for} \quad G/G/1. \tag{S.8}$$

If the expected remaining service time is non-increasing, or if the probability density of service termination (hazard function in life-testing; age-specific mortality in life insurance) is non-decreasing then certainly $R \leqslant 1/\mu$.

From (S.7) and (S.8) it follows that for $G/G/1$

$$l/\rho - 1 \leqslant l/\rho - 1 + Q_0 \leqslant q \leqslant l/\rho. \tag{S.9}$$

Section 3

The Pollaczek-Khintchine relation can be extended to the system $M/G/1$ but limited to N customers, i.e. arrival rate is λ when $n < N$ and is zero while $n = N$, when the system is full.

(S.1) is still valid but with (cf. theorem D)

$$q = \frac{l}{1 - P_N} \quad \text{and} \quad Q_0 = \frac{P_0}{1 - P_N} \tag{S.10}$$

and becomes

$$w = \frac{q}{1 - P_N} + \frac{1 - P_0 - P_N}{1 - P_N} R. \tag{S.11}$$

The effective, i.e. the « successful » arrival rate is now

$$\lambda_s = \lambda(1 - P_N) \tag{S.12}$$

and (S.2) becomes

$$l = \lambda_s w = \lambda(1 - P_N) w.$$

From (S.11) and (S.12) we get

$$l = \frac{\lambda R}{1-\rho}(1 - P_0 - P_N). \tag{S.13}$$

The conservation of customers (ref. 1) requires that

$$\lambda(1 - P_N) = \mu(1 - P_0), \tag{S.14}$$

i.e. that the (effective) input frequency equals the (effective) output frequency.

Therefore (S.13) can be written as

$$l = \frac{\lambda R}{1-\rho}\frac{1}{\rho}[1 - (1 + \rho)P_0]. \tag{S.15}$$

P_0 has to be derived by other means, so that (S.10) and (S.15) are incomplete.

Of course, the factor R is the same in the bounded as in the unbounded queuing system since it refers to the full-load operation. The linear forms (S.13) and (S.15) can be transformed into quadratic forms using (S.6).

Observe that (S.13) becomes (S.5) when N tends to infinity, as expected, since P_N tends then to zero.

When $N = 1$ no waiting line for service is allowed and we have $P_0 + P_1 = 1$; our formula (S.13) yields, of course, $l = 0$ in this case.

Section 4

The Pollaczek-Khintchine formula (S.5) can be extended to the case when the customers arrive in groups, the events being distributed in time in a Poissonian fashion. The size of the group can itself be a random variable.

Let

(S.16) a_i = probability that an arriving group has i members; $i \geqslant 1$.

(S.17) $A = E(a_i) = \Sigma i a_i$

(S.18) λ_g = frequency of group arrivals

(S.19) $\lambda = A\lambda_g$ = frequency of customer arrivals.

Then the expected waiting time for service is

$$w = q\frac{1}{\mu} + (1 - Q_0)R + G = l/\lambda \tag{S.20}$$

where

(S.21) G = average time a member of an arriving group waits for his fellows to be serviced once the « old » customers have left the queuing system.

If there are k customers in a group of arrivals then the average time of waiting for service after the « old » customers have left the system is the average of $\frac{1}{\mu} + \frac{2}{\mu} + \frac{k-1}{\mu}$, i.e. $G_K = \frac{1}{\mu}\frac{k-1}{2}$.

In turn, the average of the G_k is $G = \sum_1^\infty a_k G_k = \frac{1}{2\mu}(A-1)$.

According to the Scholion to theorem B, $q = l$ and, of course, we have $Q_0 = P_0$ (probability that the booth is empty) for $M/G/1$, with or without bulking. Therefore

(S.22)
$$w = l\frac{1}{\mu} + \rho R + \frac{A-1}{2\mu} = l/\mu$$

and

(S.23)
$$l = \frac{\lambda\rho R}{1-\rho} + \frac{\rho}{1-\rho}\frac{A-1}{2}.$$

(S.23) is the result sought for bulked arrivals. Note that when $A = 1$ we get the usual Pollaczeck-Khintchine formula (S.5) in linear form.

The fact that the queue length l increases with growing average bulk size A is plausible. But the fact that the loss is proportional to $A - 1$ is not intuitively obvious.

BIBLIOGRAPHY

[1] M. KRAKOWSKI, *Conservation Methods in Queuing Theory*, R.A.I.R.O., 7e année, V-1, 1973, p. 63-83.

[2] LITTLE, John, D. C., *A Proof for the Queuing Formula : $L = \lambda W$*, Oper. Res., vol. 9, No 3, 383-387, 1961.

[3] D. R. COX and W. L. SMITH, *Queues*, John Wiley & Sons, 1961.

[4] SAATY, THOMAS L., *Elements of Queuing Theory with Applications*, McGraw-Hill, 1961.