COHEN, J. W. *The Single Server Queue*. North-Holland, Amsterdam; Wiley Interscience, New York, 1969. xiv + 657 pp. $35.00.

Review by LAJOS TÁKACS

Case Western Reserve University

The theory of queues deals with the mathematical studies of random mass service phenomena. Such phenomena appear in physics, engineering, industry, transportation, commerce, business, and several other fields. Many processes arising in these fields can be described by the following queuing model: In the time interval $(0, \infty)$ customers arrive at a counter at random times, $\tau_1, \tau_2, \cdots, \tau_n, \cdots$, and are served by one or more servers. The successive service times, $\chi_1, \chi_2, \cdots, \chi_n, \cdots$, are random variables. The initial state is determined by the initial queue size and by the initial occupation times of the servers.

The most important problems are connected with the investigation of the stochastic behavior of the waiting time, the queue size, the busy periods, the idle periods, and the departures. It is of interest to find the distributions of the actual waiting time of the nth customer, the virtual waiting time at time t (the time which a customer would have to wait if he arrived at time t), the queue size immediately before the arrival of the nth customer, the queue size at time t, the queue size immediately after the nth departure, and the lengths of the busy periods and the idle periods. (Busy periods and idle periods are successive time intervals during which there is at least one customer in the system or there is no customer in the system.)

One of the most important mathematical models is the following: The inter-arrival times, $\{\tau_n - \tau_{n-1}\}$, and the service times, $\{\chi_n\}$, are independent sequences of mutually independent and identically distributed positive random variables, and the customers are served by m servers in the order of arrival. Let $\mathbf{P}\{\tau_n - \tau_{n-1} \leqq x\} = F(x)$ and $\mathbf{P}\{\chi_n \leqq x\} = H(x)$.

This model in the particular case when $F(x) = 1 - e^{-\lambda x}$ for $x \geqq 0$, $H(x) = 1$ for $x \geqq \alpha$, $H(x) = 0$ for $x < \alpha$, and $m = 1$ was studied for the first time in 1909 by A. K. Erlang who showed that if $\lambda\alpha < 1$, then the waiting time η_n has a stationary distribution, and if $\mathbf{P}[\eta_n \leqq x] = W(x)$, then

$$W(x) = (1-\lambda\alpha)\sum_{j=0}^{[x/\alpha]}(-1)^j e^{\lambda(x-j\alpha)}\frac{[\lambda(x-j\alpha)]^j}{j!}$$

for $x \geqq 0$.

In 1930, F. Pollaczek, and 1932, A. Y. Khintchine, showed that if $m = 1$, $F(x) = 1 - e^{-\lambda x}$ for $x \geqq 0$, $H(x)$ has a finite expectation α, and $\lambda\alpha < 1$, then the waiting time η_n has a limiting distribution function $W(x)$ which is independent of the initial state, and $W(x)$ is the unique stationary distribution. They found that

$$\int_0^\infty e^{-sx}\,dW(x) = \frac{1-\lambda\alpha}{1-\lambda\dfrac{1-\psi(s)}{s}}$$

for $\operatorname{Re}(s) > 0$ where $\psi(s) = \int_0^\infty e^{-sx}\,dH(x)$. Hence it follows that the expectation of the stationary distribution of the waiting time is

$$\int_0^\infty x\,dW(x) = \frac{\lambda(\sigma^2+\alpha^2)}{2(1-\lambda\alpha)}$$

where σ^2 is the variance of the service times. The last formula shows an interesting peculiarity, namely, that the average waiting time depends not only on the average interarrival times, $1/\lambda$, and the average service times, α, but it depends on the variance of the service times too. This is a typical phenomenon in queuing theory.

These classical results have largely been extended in the past forty years. Various service disciplines have been introduced: first-come first-served, last-come first-served, random service, priority service, service in batches. Different arrival processes have been considered: Poisson input, recurrent input, superposition of various input processes, group arrivals. Queuing networks have been investigated: queues in series, queues in parallel. Queues with limited queue size and limited waiting time (waiting room etc.) have been studied. Attention has been shifted from the stationary behavior of queues to the time-dependent behavior. New mathematical methods have been developed for studying various queuing models.

In this book the author gives a general survey of numerous results for single-server queues with various queuing disciplines. The main emphasis is on the time-dependent behavior of such queues. The author himself contributed significantly to these studies and the book contains many of his own results which were available only in scientific papers until now.

Part I deals with various topics in stochastic processes which have importance in queuing theory. In particular, Markov chains, Markov processes, derived Markov chains, renewal processes, and fluctuation theory are discussed.

Part II deals with the classical model of single-server queues in various particular cases. The cases of Poisson input, Erlang input, recurrent input, and exponential, Erlang, and general service times are discussed with great thoroughness.

Part III deals with various generalizations of single-server queues. The author studies bulk queues, in which customers are arriving in groups and are served in batches, priority queues, and finite queues, in which the queue size or the waiting time is limited. Finite queues have importance in the theories of storage and dams too. The book ends with a chapter on limit theorems for single-server queues.

The book covers many important mathematical models in the theory of single-server queues, and in each case the author provides an extensive and thorough study. Many of the results were achieved by the author in the past several years. The book is well organized and it is a useful source of a great many important results. The author uses mainly a method of Pollaczek. Pollaczek's method, which is based on complex integrals, has the obvious advantage that it is unified and provides the solution in a closed form for a large number of queuing processes. It should be added, however, that if we apply Pollaczek's method in queuing theory, then we should impose some restrictions either on the distribution of the inter-arrival times or on the distribution of the service times.

Finally, it is relevant to mention briefly the essence of Pollaczek's method. We shall present this method here in a more general form than the original one. This generalization makes it possible to avoid imposing any unnecessary conditions on the distributions in question.

Let R denote the space of functions $\Phi(s)$ defined for $\mathrm{Re}(s) = 0$ on the complex plane, which can be represented in the form

$$\Phi(s) = \mathbf{E}\{\zeta e^{-s\eta}\}$$

where ζ is a complex (or real) random variable with $\mathbf{E}\{|\zeta|\} < \infty$ and η is a real random variable. Define a linear transformation $\mathbf{T}$ for $\Phi(s) \in R$ such that

$$\mathbf{T}\Phi(s) = \Phi^+(s) = \mathbf{E}\{\zeta e^{-s\eta^+}\}$$

for $\mathrm{Re}(s) \geqq 0$ where $\eta^+ = \max(0, \eta)$.

If $\Phi(s) \in R$, then

$$\Phi^+(s) = \tfrac{1}{2}\Phi(0) + \lim_{\varepsilon\to 0}\frac{s}{2\pi i}\int_{L_\varepsilon}\frac{\Phi(z)}{z(s-z)}dz$$

for $\mathrm{Re}(s) > 0$ where $L_\varepsilon = \{z = iy : -\infty < y < -\varepsilon < \varepsilon < y < \infty\}$.

Many problems in the theory of queues can be formulated in the following way: Let $\gamma(s) \in R$, $\Gamma_0(s) \in R$ and $\mathbf{T}\Gamma_0(s) = \Gamma_0(s)$ and define $\Gamma_n(s)$ $(n = 1, 2, \cdots)$ by the recurrence relation

$$\Gamma_n(s) = \mathbf{T}\{\gamma(s)\Gamma_{n-1}(s)\}$$

for $n = 1, 2, \cdots$. The functions $\Gamma_n(s)$ $(n = 1, 2, \cdots)$ are to be determined.

We have

$$\sum_{n=0}^{\infty}\Gamma_n(s)\rho^n = \exp[-\mathbf{T}\{\log[1-\rho\gamma(s)]\}]\,\mathbf{T}\left\{\frac{\Gamma_0(s)\exp[\mathbf{T}\{\log[1-\rho\gamma(s)]\}]}{1-\rho\gamma(s)}\right\}$$

for sufficiently small $|\rho|$. If, in particular, $\Gamma_0(s) = 1$, then

$$\textstyle\sum_{n=0}^{\infty}\Gamma_n(s)\rho^n = \exp[-\mathbf{T}\{\log[1-\rho\gamma(s)]\}]$$

for sufficiently small $|\rho|$.

In his studies Pollaczek considered a smaller space than R, in this space expressed $\Phi^+(s)$ with the aid of a simple Cauchy integral and obtained $\sum_{n=0}^{\infty}\Gamma_n(s)\rho^n$ as a solution of an integral equation.