Z. Wahrscheinlichkeitstheorie verw. Gebiete
51, 115–121 (1980)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

On the Accuracy of Kingman's Heavy Traffic Approximation in the Theory of Queues

Walter A. Rosenkrantz*

Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003, USA

I. Introduction

We consider a family of $GI/GI/1$ queues $Q(\alpha)$, depending on a parameter α, and set $W(n,\alpha)$ = waiting time, excluding service, of the n^{th} customer, $S(n,\alpha)$ = service time of the n^{th} customer and $T(n,\alpha)$ = time between arrival of n^{th} and $(n+1)$st customer. The n^{th} customer is denoted by $C(n)$ and we assume $C(0)$ arrives at time $t=0$ and finds the server free. Let $U(n,\alpha)=S(n,\alpha)-T(n,\alpha)$ and assume

$$E(U(n,\alpha))=-\alpha\sigma, \qquad V(U(n,\alpha))=\sigma^2, \quad \alpha>0. \tag{1}$$

The "heavy traffic approximation" of Kingman - see [2, 3] - is a limit theorem for $(\alpha/\sigma)\,W(n,\alpha)$ as $\alpha\to 0$, $n\to\infty$ in such a way that $\lim_{n\to\infty,\alpha\to 0}\alpha^2 n=+\infty$. More precisely we have

Theorem 1 (Kingman). $\lim_{n\to\infty,\alpha\to 0} P((\alpha/\sigma)\,W(n,\alpha)\leqq x)=1-e^{-2x}$, $0\leqq x<\infty$, *provided* $\lim_{n\to\infty,\alpha\to 0}\alpha^2 n=\infty$.

Our main purpose is to present in Theorem 2 below an estimate of the error in Theorem 1 which to the best of our knowledge is new and under certain hypotheses nearly optimal. Let $F_{n,\alpha}(x)=P((\alpha/\sigma)\,W(n,\alpha)\leqq x)$ and $F(x)=1-e^{-2x}$ and denote the error by $e(n,\alpha)=\operatorname{Sup}_{0\leqq x<\infty}|F_{n,\alpha}(x)-F(x)|$. To simplify matters let us first assume that the normalized random variables $X_n=(U(n,\alpha)+\alpha\sigma)/\sigma$ do not depend on α, thus X_n form an i.i.d. sequence with $E(X_n)=0$, $V(X_n)=+1$. The order of magnitude of the error $e(n,\alpha)$ as a function of both α and n depend on the hypotheses we impose on X_n. The following conditions indicate some of the possibilities:

$$R(t)=E(e^{tX_n}) \quad \text{is convergent for } |t|<t_0. \tag{2}$$

Remark. This is the hypothesis of Theorem 1 of Komlos-Major-Tusnady [4] which we use later.

$$E(|X_n|^p)<\infty, \qquad p>2. \tag{3}$$

* This research was supported by the Air Force office of scientific research contract F 49620-79-C-0209

0044-3719/80/0051/0115/$01.40

Theorem 2. (i) *Assume the common moment generating function $R(t)$ of the random variables X_n satisfies condition* (2) *and $\alpha^2 n = 2\log n$, so $\alpha = (2\log n/n)^{1/2}$, then*

$$e(n,\alpha) \leqq K(\log n)^{3/2}\, n^{-1/2}$$

where K depends only on the distribution of X_n.

(ii) *Assume condition* (3) *and $\alpha^2 n = 2\log n$, then*

$$e(n,\alpha) \leqq K'(\log n)^{(p/2p+2)} \cdot n^{(2-p)/(2p+2)}, \qquad \textit{all } p > 2.$$

Remarks. (i) If we allow X_n to depend on α in such a way that either of the conditions (2) and (3) hold uniformly in α then the conclusion of Theorem 2 remains valid.

(ii) Our original version of Theorem 2 (ii) relied on the Skorokhod embedding theorem as in the author's papers [6] and [7] and yielded a rate of convergence that was at best $(\log n)\, n^{-1/4}$, cf. Dudley [11], Theorem 5.2. The much improved version given here as well as the proof i.e. Theorem 4 are due to S. Csörgö [10] to whom the author is greatly indebted for his most valuable suggestions.

(iii) The error estimate of Theorem 2 allows one to deduce a sort of "large deviations" theorem of which the following is typical:

$$\begin{aligned} P(W(n,\alpha) \geqq x_n) &= P((\alpha/\sigma)\, W(n,\alpha) \geqq (\alpha/\sigma)\, x_n)) \\ &= e^{-(2\alpha x_n/\sigma)} \pm e(n,\alpha). \end{aligned}$$

If $\lim\limits_{n\to\infty,\,\alpha\to 0} e(n,\alpha)\, e^{(2\alpha x_n/\sigma)} = 0$ then we have

$$1 - F_{n,\alpha}(\alpha x_n/\sigma)) \sim e^{-2\alpha x_n/\sigma}, \qquad \text{as } \alpha \to 0,\ n \to \infty$$

in such a way that $\alpha^2 n \to +\infty$.

This may be regarded as a refinement of the inequalities (24), (25) on p. 146 of [3].

For example if we replace the choice of $\alpha^2 n = 2\log n$ by $\alpha^2 n = 2a_n$ where $a_n \uparrow +\infty$ at a certain rate then our proof yields

Theorem 2′. (i) *If $R(t) < \infty$, $|t| < t_0$ then*

$$e(n,\alpha) = O(\max(a_n^{1/2}(\log n/n)^{1/2}, e^{-a_n}/a_n^{1/2}))$$

provided $a_n = O(n/\log^2 n)$ as $n \uparrow +\infty$.

(ii) *If $E(|X_n|^p) < \infty$, $p > 2$ then*

$$e(n,\alpha) = O((a_n)^{p/(2p+2)}/n^{(p-2)/(2p+2)})$$

provided $a_n < n^{(p-2)/p}$.

We omit the details of the proof as all one has to do is retrace the steps of the proof of Theorem 2 replacing $\log n$ by a_n.

II. Proof of Main Theorem

Our first step is to represent $(\alpha/\sigma)\,W(n,\alpha)$ as a functional of a stochastic process $y_{n,\alpha}(t)$ which converges to the Wiener process with drift, denoted by $x(t)=w(t)-t$. To this end we note that

$$W(n,\alpha)=\operatorname*{Max}_{0\leqq k\leqq n} Y_k(n,\alpha) \quad \text{where } Y_0(n,\alpha)\equiv 0 \quad \text{and for } k\geqq 1 \tag{4}$$

$$Y_k(n,\alpha)=\sum_{i=1}^{k} U(i,\alpha)=\sigma\sum_{i=1}^{k} X_i-k\,\alpha\,\sigma$$

and thus

$$(\alpha/\sigma)\,W(n,\alpha)=\operatorname*{Max}_{0\leqq k\leqq n}(S_k-k\,\alpha^2) \quad \text{where } S_0=0 \quad \text{and} \quad S_k=\sum_{i=1}^{n}\alpha X_i. \tag{5}$$

Following Prohorov [5] let us define the "random broken line" $y_{n,\alpha}(t)=S_k-k\,\alpha^2$ for $t=k\alpha^2$, $0\leqq k\leqq n$ and define it for values $k\alpha^2\leqq t\leqq(k+1)\,\alpha^2$ by linear inter polation. $y_{n,\alpha}(t)$ is thus a stochastic process with continuous paths whose local maxima are attained at the nodes $k\alpha^2$. Hence

$$F_{n,\alpha}(x)=P(\operatorname*{Max}_{0\leqq t\leqq\alpha^2 n} y_{n,\alpha}(t)\leqq x) \text{ - see pp. 146–147 of [3].} \tag{6}$$

It is natural to first estimate $|F_{n,\alpha}(x)-\tilde F_{n,\alpha}(x)|$ where

$$\tilde F_{n,\alpha}(x)=P(\operatorname*{Sup}_{0\leqq t\leqq n\alpha^2} x(t)\leqq x) \tag{7}$$

and then estimate $|\tilde F_{n,\alpha}(x)-F(x)|$ where

$$F(x)=P(\operatorname*{Sup}_{0\leqq t<\infty} x(t)\leqq x)=P(\operatorname*{Sup}_{0\leqq t<\infty} w(t)-t\leqq x)=1-e^{-2x}, \tag{8}$$

see Karlin-Taylor [1], p. 361 for a proof of (8).

To estimate $F_{n,\alpha}(x)-\tilde F_{n,\alpha}(x)$ via the Skorohod embedding as in [6], [7] requires us to define the "random broken line" $x_\alpha(t)$ obtained by putting $x_\alpha(t)=x(t)$ if $t=k\alpha^2$ and define $x_\alpha(t)$ for other values of t by linear interpolation. Clearly $x_\alpha(t)=w_\alpha(t)-t$, where $w_\alpha(t)$ is obtained from the Wiener process by linearly interpolation at $t=k\alpha^2$, $k=0,1,\ldots,n$. Our problem now is to estimate $e_1(n,\alpha)$, $e_2(n,\alpha)$ where

$$e_1(n,\alpha)=\operatorname*{Sup}_{0\leqq x<\infty}|F_{n,\alpha}(x)-\tilde F_{n,\alpha}(x)|, \tag{9}$$

$$e_2(n,\alpha)=\operatorname*{Sup}_{0\leqq x<\infty}|\tilde F_{n,\alpha}(x)-F(x)| \tag{10}$$

and noting that $e(n,\alpha)\leqq e_1(n,\alpha)+e_2(n,\alpha)$. It is easiest to begin by obtaining an estimate for $e_2(n,\alpha)$. Let $\Phi(x)=\int_{-\infty}^{x}(2\pi)^{-1/2}e^{-y^2/2}\,dy$ and recall the tail estimate.

$$1-\Phi(x)<c\,x^{-1}e^{-x^2/2}, \quad \text{as } x\to\infty,\ c=(2\pi)^{-1/2}. \tag{11}$$

Shepp, [8], p. 348, has shown that

$$P(w(t)<at+b, 0\leqq t\leqq L)=\Phi\left(\frac{aL+b}{L^{1/2}}\right)-e^{-2ab}\,\Phi\left(\frac{aL-b}{L^{1/2}}\right) \tag{12}$$

and this clearly implies that

$$\tilde{F}_{n,\alpha}(x)=\Phi\left(\frac{\alpha^2 n+x}{\alpha n^{1/2}}\right)-e^{-2x}\,\Phi\left(\frac{\alpha^2 n-x}{\alpha n^{1/2}}\right) \tag{13}$$

and thus

$$e_2(n,\alpha)\leqq \sup_{0\leqq x<\infty}\left|\Phi\left(\frac{\alpha^2 n+x}{\alpha n^{1/2}}\right)-1\right|+\sup_{0\leqq x<\infty}\left|\left(\Phi\left(\frac{\alpha^2 n-x}{\alpha n^{1/2}}\right)-1\right)e^{-2x}\right|. \tag{14}$$

Now $\Phi\left(\frac{\alpha^2 n+x}{\alpha n^{1/2}}\right)\geqq\Phi(\alpha n^{1/2})$ and so

$$\left|1-\Phi\left(\frac{\alpha^2 n+x}{\alpha n^{1/2}}\right)\right|\leqq|1-\Phi(\alpha n^{1/2})|\leqq(c/\alpha n^{1/2})\,e^{-\frac{\alpha^2 n}{2}},$$

where we've used the tail estimate (11).

If $\alpha^2 n=2\log n$ then clearly $\left|1-\Phi\left(\frac{\alpha^2 n+x}{\alpha n^{1/2}}\right)\right|\leqq c\cdot 2^{-1/2}n^{-1}(\log n)^{-1/2}$. To estimate the second summand in (14) we calculate the sup over the intervals $0\leqq x\leqq\alpha^2 n/2$ and $(\alpha^2 n/2)\leqq x<\infty$. On the interval $[0,\alpha^2 n/2]$ $\frac{\alpha^2 n-x}{\alpha n^{1/2}}\geqq\frac{1}{2}\alpha\sqrt{n}$ and thus

$$\left|\left(\Phi\left(\frac{\alpha^2 n-x}{\alpha n^{1/2}}\right)-1\right)e^{-2x}\right|\leqq\frac{c}{\alpha n^{1/2}}\,e^{-\alpha^2 n/2}=c\cdot 2^{-1/2}n^{-1}(\log n)^{-1/2}.$$

If $x\geqq(\alpha^2 n/2)$ then $e^{-2x}\leqq e^{-\alpha^2 n}=e^{-2\log n}=n^{-2}$ and so

$$e_2(n,\alpha)\leqq n^{-1}(\log n)^{-1/2}\quad\text{if } \alpha^2 n=2\log n. \tag{15}$$

To estimate $e_1(n,\alpha)$ we proceed by first estimating

$$P(\sup_{0\leqq t\leqq n\alpha^2}|x_\alpha(t)-x(t)|>\varepsilon)=P(\sup_{0\leqq t\leqq n\alpha^2}|(w_\alpha(t)-w(t))|>\varepsilon).$$

It is not too difficult to show that

$$P(\sup_{0\leqq t\leqq n\alpha^2}|(w_\alpha(t)-w(t)|>\varepsilon)<2n\,e^{-2(\varepsilon/\alpha)^2}, \tag{16}$$

see, for example, Lemma 3, p. 546 of [6].

At this point we use an important refinement of the Skorohod embedding scheme due to Komlos-Major-Tusnady [4] to estimate $P(\sup_{0\leqq t\leqq n\alpha^2}|y_{n,\alpha}(t)-x_\alpha(t)|\geqq\varepsilon)$.

Theorem 3 (Komlos-Major-Tusnady). *Given any sequence of i.i.d. random variables $X_1, \ldots, X_n$ satisfying $E(X_i)=0$, $V(X_i)=1$ and condition (2) there exists an i.i.d. sequence of standard normal random variables $Z_1, \ldots, Z_n$, which we may assume to be defined on the same probability space, such that*

$$P\Big(\operatorname*{Max}_{1 \leq k \leq n}\Big|\Big(\sum_{i=1}^{k} X_i\Big)-\Big(\sum_{i=1}^{k} Z_i\Big)\Big|>C\log n+x\Big)<K e^{-\lambda x} \tag{17}$$

where C, K, λ depend only on the distribution of the X_i's.

This theorem permits us to assume that the "random broken lines" $y_{n,\alpha}(t)$ and $X_\alpha(t)$ are defined on the same probability space; more precisely, if we set $S_k = \sum_{i=1}^{k} \alpha Y_i$ and $T_k = \sum_{i=1}^{k} \alpha Z_i$, then $X_\alpha(t)$ is obviously identical in law to the "random broken line" $\hat{X}_\alpha(t)=T_k-k\alpha^2$ at $t=k\alpha^2$ and defining $\hat{X}_\alpha(t)$ for other values of t by linear interpolation. Hence

$$\begin{aligned}
&P(\operatorname*{Sup}_{0 \leq t \leq n\alpha^2} |y_{n,\alpha}(t)-\hat{X}_\alpha(t)|>\alpha(C\log n+x)) \\
&\quad =P(\operatorname*{Max}_{1 \leq k \leq n} |S_k-T_k|>\alpha(C\log n+x)) \\
&\quad =P\Big(\operatorname*{Max}_{1 \leq k \leq n}\Big|\Big(\sum_{i=1}^{k} X_i\Big)-\Big(\sum_{i=1}^{k} Z_i\Big)\Big|>C\log n+x\Big)<K e^{-\lambda x}.
\end{aligned}$$

By an abuse of notation we write

$$P(\operatorname*{Sup}_{0 \leq t \leq n\alpha^2} |y_{n,\alpha}(t)-x_\alpha(t)|>\alpha(C\log n+x)<K e^{-\lambda x}. \tag{18}$$

Set $\hat{F}_{n,\alpha}(x)=P(\operatorname*{Sup}_{0 \leq t \leq n\alpha^2} x_\alpha(t)<x)$, let $x=\beta \log n$ in (18) and put $\varepsilon(n,\alpha)=\alpha(C+\beta)\log n$. Then from (18) we get

$$P(\operatorname*{Sup}_{0 \leq t \leq n\alpha^2} |y_{n,\alpha}(t)-X_\alpha(t)| \geqq \varepsilon(n,\alpha)) \leqq K n^{-\lambda\beta} \tag{19}$$

where β can be chosen to be arbitrarily large. An immediate consequence of (19) is the estimate

$$\hat{F}_{n,\alpha}(x-\varepsilon(n,\alpha))-K n^{-\lambda\beta} \leqq F_{n,\alpha}(x) \leqq \hat{F}_{n,\alpha}(x+\varepsilon(n,\alpha))+K n^{-\lambda\beta}. \tag{20}$$

Similarly estimate (16) yields the inequalities

$$\tilde{F}_{n,\alpha}(x+\varepsilon)-2n e^{-2(\varepsilon/\alpha)^2} \leqq \hat{F}_{n,\alpha}(x) \leqq \tilde{F}_{n,\alpha}(x-\varepsilon)+2n e^{-2(\varepsilon/\alpha)^2}.$$

Setting $\alpha^2 n=2\log n$ and combining estimates (15), (16), (19) and (20) leads to the result that

$$\begin{aligned}
&|F_{n,\alpha}(x)-F(x)| \leqq 2^{3/2}(C+\beta)(\log n)^{3/2} n^{-1/2}+K n^{-\lambda\beta} \\
&+2n e^{-2(C+\beta)(\log n)^2}+e_2(n,\alpha),
\end{aligned}$$

where the right hand side is independent of x and is clearly of $O(\log n)^{3/2} n^{-1/2})$ order of magnitude, provided β is chosen large enough.

To prove Theorem 2(ii) we make use of the following Theorem which is a consequence of Theorem 4 of [4] and Theorem 2 of [9].

Theorem 4. *Let $H(x)>0$, $x>0$ be a non decreasing continuous function such that*

(i) *$H(x)/x^{2+\varepsilon}$ is non decreasing for some $\varepsilon>0$.*

(ii) *$\log H(x)/x$ is non decreasing and*

(iii) *$E(|H(X_i)|)<\infty$ where the X_i are an i.i.d. sequence with $E(X_i)=0$, $E(X_i^2)=1$,*

Then there exists an i.i.d. sequence of standard normal random variables $Z_1, \ldots, Z_n$ which we may assume to be defined on the same probability space, such that

$$P\left(\operatorname*{Max}_{1\leqq k\leqq n}\left|\left(\sum_{i=1}^{k} X_i\right)-\left(\sum_{i=1}^{k} Z_i\right)\right|>x_n\right)<c_2\, n/H(c_3\, x_n), \tag{22}$$

for all x_n such that $H^{-1}(n)<x_n<c_1\sqrt{n\log n}$ where c_1, c_2, c_3 depend only on the underlying distribution of the $\{X_i\}$.

To apply this result to Theorem 2(ii) we set $H(x)=x^p, p>2$ with $H^{-1}(x)=x^{1/p}$ and x_n chosen so that $n^{1/p}<x_n<c_1(n\log n)$. With this choice it is easy to show that $c_2\, n/H(c_3\, x_n)<k\, n\, x_n^{-p}$ and in particular if one chooses $x_n=n^{3/(2+2p)}/(\log n)^{1/(2+2p)}$. Then estimate (18) becomes

$$P(\operatorname*{Sup}_{0\leqq t\leqq n\alpha^2}|y_{n,\alpha}(t)\quad x_\alpha(t)|>\varepsilon_n)<k(\log n)^{p(2+2p)}/n^{(p-2)/(2p+2)} \tag{23}$$

where $\varepsilon_n=(\log n)^{p/(2+2p)}\, n^{(1-p)/(2+2p)}, \quad p>2.$

Concluding Remarks. The referee has inquired if anything can be said about the size of the constants K, K' of Theorem 2. If the original Skorohod embedding were used then Theorem 2(ii) in the case $p=4$ would become $e(n,\alpha)\leqq K(\log n)^{1/2} n^{-1/5}$ where $K\leqq 35$. However the rate of convergence as $p\to\infty$ is no better than $n^{-1/4}$, up to logarithmic terms via this method. Note that the asymptotic rate via the $K-M-T$ method is $n^{-1/2}$, up to logarithmic terms. Unfortunately the size of the constants C, K, λ in Theorem 3 are at present unknown, according to Csörgö [10].

References

1. Karlin, H., Taylor, S.J.: A First Course in Stochastic Processes. New York: Academic Press 1975
2. Kingman, J.F.C.: On Queues in Heavy Traffic. J. Roy. Statist. Soc., ser. B, **24**, 383–392 (1962)
3. Kingman, J.F.C.: The Heavy Traffic Approximation in the Theory of Queues. Proc. Sympos. Congestion Theory. Chapel Hill: University of North Carolina Press, pp. 137–169, 1965
4. Komlos, J., Major, P., Tusnady, G.: An approximation of partial sums of independent random variables and the sample DF. II. Z. Wahrscheinlichkeitstheorie verw. Gebiete **34**, 33–58 (1976)

5. Prohorov, Y.: Convergence of random Processes and limit theorems in Probability. Theor. Probability Appl. **1**, 157–214 (1956)
6. Rosenkrantz, W.: On rates of convergence for the invariance principle. Trans. Amer. Math. Soc. **129**, 542–552 (1967)
7. Rosenkrantz, W.: A rate of convergence for the Von Mises Statistic. Trans. Amer. Math. Soc. **139**, 329–337 (1969)
8. Shepp, L.: Radon-Nikodym Derivatives of Gaussian Measures. Ann. Math. Statist. **37**, 321–354 (1966)
9. Major, P.: The Approximation of Partial Sums of Independent RV's. Z. Wahrscheinlichkeitstheorie verw. Gebiete **35**, 213–220 (1976)
10. Csörgö, S.: (Private communication, 5/24/79).
11. Dudley, R.M.: Speeds of Metric Probability Convergence. Z. Wahrscheinlichkeitstheorie **22**, 323–332 (1972)

Received May 1, 1979