A Central Limit Theorem for Interval Packing

E. G. Coffman Jr. Leopold Flatto Predrag Jelenković

Bell Labs
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974
{egc, leopold, predrag}@bell-labs.com

Jan. 19, 1998

Abstract

Starting at time 0, unit-length intervals arrive and are placed on the positive real line by a unit-intensity Poisson process in two dimensions; the probability of an interval arriving in $[t, t+dt]$ with its left endpoint in $[y, y+dy]$ is $dtdy + o(dtdy)$. An arrival is *accepted* if and only if, for some given x, the interval is contained in $[0, x]$ and overlaps no interval already accepted.

We study the number $N(t,x)$ of intervals accepted during $[0,t]$. Using known asymptotic estimates for the first two moments of $N(t,x)$, we show that $N(t,x)$ satisfies a central limit theorem (CLT), i.e., for any fixed t

$$\frac{N(t,x) - \mathrm{E}N(t,x)}{\sigma(N(t,x))} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad x \to \infty,$$

where $\mathcal{N}(0,1)$ is a standard normal random variable, and $\xrightarrow{d}$ denotes convergence in distribution. This stochastic, on-line interval packing problem generalizes the classical parking problem, the latter corresponding only to the absorbing states of the interval packing process, where successive packed intervals are separated by gaps less than 1 in length. Coffman et al [1] sketched a proof of the above CLT, pointing out that the proof followed closely the argument used by Dvoretzky and Robbins [2] to prove the corresponding CLT for the parking problem. Here, we give all the details, and show exactly where and how the steps in the Dvoretzky-Robbins proof have to be modified in a proof of the CLT for the more general stochastic interval packing problem.

1 Introduction

Unit intervals arrive at random times and at random locations in $\mathbf{R}_+$. The arrival times and left endpoints comprise a unit-intensity Poisson process in two dimensions; the probability of an arrival in the time interval $[t, t+dt]$ with left endpoint in $[y, y+dy]$ is $dtdy + o(dtdy)$. The number packed during $[0,t]$ in $[0,x]$ is denoted by $N(t,x)$ In [1], it is shown that, for any given $T > 0$,

$$\sup_{0\le t\le T} |\mathrm{E}N(t,x) - (\alpha(t)x + \alpha(t) + \beta(t) - 1)| = O\left(e^{-\xi x \log x}\right) \tag{1}$$

for all $\xi \in (0,1)$, where

$$\alpha(t) = e^{\beta(t)}, \quad \beta(t) = -2\int_0^t \frac{1-e^{-u}}{u} du. \tag{2}$$

And for the variance, it is shown in [1] that, for any given $T > 0$,

$$\sup_{0\le t\le T} \left|\sigma^2(N(t,x)) - (\mu(t)x + \mu_1(t))\right| = O(e^{-\xi x \log x}), \tag{3}$$

for all $\xi \in (0,1)$, where $\mu(t)$ and $\mu_1(t)$ are explicitly computable (see [1, eqs. (75) and (76)]), and where $\mu(t) > 0$ for all $t > 0$.

The classical parking problem of Renyi [4] has an intimate relationship with on-line interval packing. In the former problem, unit-length cars are parked sequentially along a curb (interval) $[0,x]$, $x > 1$. Each car chooses a parking place independently and uniformly at random from those available, i.e., from those where it will not overlap cars already parked or the curb boundaries. This uniform parking of cars continues until every unoccupied gap is less than 1 in length, i.e., no further cars can be parked. It is verified in [1] that, as might be expected, $N(t,x)$ tends to $\tilde{N}(x)$ in distribution as $t \to \infty$, where $\tilde{N}(x)$ is the number parked at the conclusion of the parking process. Extending results of Renyi [4], Dvoretzky and Robbins [2] showed that the mean of $\tilde{N}_x$ has the estimate in (1), once the limit $t \to \infty$ is taken, where $\alpha(\infty) = 0.748\ldots$. Similarly, the combined results of Dvoretzky and Robbins [2] and Mackenzie [3] showed that the variance of $\tilde{N}_x$ has the estimate in (3), once the limit $t \to \infty$ is taken; in this limit

$$\begin{aligned} \mu(\infty) = \mu_1(\infty) &= 4\int_0^\infty \left[e^{-y}(1-e^{-y})\frac{\tilde{\alpha}(y)}{y} - \frac{e^{-2y}(e^{-y}-1+y)\tilde{\alpha}^2(y)}{\beta(y)y^2}\right] dy - \alpha(\infty), && (4)\\ &= 0.0381\ldots, && (5) \end{aligned}$$

where $\tilde{\alpha}(y) := \alpha(\infty) - \alpha(y)$. For additional literature on the parking problem, see [1]

Although one expects some strong form of convergence of $N(t,x)$ to $\tilde{N}(x)$, it is surprising at first to find that, for all $x > 2$, the expected time to absorption of the interval packing process is infinite. It is shown in [1] that, if T_x denotes the time-to-absorption of the interval packing process, then for any fixed x, $\mathbf{P}(T_x > t)$ tends to 0 like $1/t$, and hence, T_x is finite almost surely, but $\mathrm{E}T_x = \infty$.

Dvoretzky and Robbins [2] gave a central limit theorem for the parking problem, basing the second of their two proofs on various properties also shared by $N(t,x)$. This being the case, one might hope to adapt the technique to our problem so as to obtain a central limit theorem for any fixed t. This is indeed possible, and it is the purpose of this note to work out the details of this adaptation in a rigorous proof.

2 The central limit theorem

We prove a central limit theorem for a suitably normalized version of $N(t,x)$, which gives us an approximation of $N(t,x)$ for any fixed t, when x is large.

Theorem 1 *For any fixed t, we have $Z(t,x) \xrightarrow{d} \mathcal{N}(0,1)$, as $x \to \infty$, where*

$$Z(t,x) = \frac{N(t,x) - \mathrm{E}N(t,x)}{\sigma N(t,x)},$$

$\mathcal{N}(0,1)$ is a standard normal random variable, and $\xrightarrow{d}$ denotes convergence in distribution.

Remarks: The main ideas, which closely track the approach of [2], are overviewed as follows. We consider the state of the packing process after a suitably small number $n_x - 1 = o(x)$ of intervals have been packed, leaving a vector of n_x successive gaps $\mathbf{y} = (y_1, \ldots, y_{n_x})$. As in [2], the choice

$$n_x := \lfloor x \ln^2 x \rfloor$$

will serve our needs, so long as we assume $x \geq 4$, so that $n_x \geq 2$. The key observation is that the continuation of the packing process consists of n_x *independent* packing subprocesses, one taking place in each gap y_i. If τ_x is the earliest time by which the initial n_x intervals are packed, then at any time $t > \tau_x$, the n_x subprocesses define a triangular array of independent random variables $N_{y_i}(t-\tau_x)$ indexed by i, $1 \leq i \leq n_x$, and $x = n_x - 1 + \sum y_i > 1$. After a suitable normalization, we apply a version of Liapounov's theorem for triangular arrays and obtain a conditional central limit theorem for $Z(t,x)$, the normalized version of $N(t,x) = n_x - 1 + \sum N(t-\tau_x, y_i)$, given $(\tau_x, \mathbf{y})$. Finally, we extend the Dvoretzky-Robbins argument and show that the central limit theorem holds uniformly over a set of $(\tau_x, \mathbf{y})$ whose probability tends to 1 as $x \to \infty$. This extension is straightforward once it is verified that $\tau_x \xrightarrow{d} 0$ as $x \to \infty$, and that the estimate of the variance of $N_x(t)$ is uniform in t. It follows that the central limit theorem also holds for the unconditional packing process $Z(t,x)$.

The main result below is Theorem 6, which together with Lemma 4 gives the CLT quite readily. Leading up to Theorem 2, there are several preliminary lemmas; Lemmas 1 and 4 can be found in [2]; Lemma 2 is the desired "uniform-in-t" version of Lemma 1 in [2]; Lemma 3 is new and deals with the limiting behavior of τ_x for large x. ∎

The remaining program is as follows. After introducing some random variables, we will prove a sequence of four lemmas leading up to Theorem 2. We will then be in position to prove Theorem 1.

Observe that the conditional distribution of $N(t,x)$, given $(\mathbf{y},\tau)$, is identical to the distribution of $(n_x-1)+\sum_{i=1}^{n_x} N(t-\tau, y_i)$, where $N(t-\tau, y_1), \ldots, N(t-\tau, y_{n_x})$ are independent random variables. Hence, as noted above, the conditional distribution of the normalized random variables $Z(t,x)$, given $(\mathbf{y},\tau)$,is identical to the distribution of $\sum_{i=0}^{n_x} Y_i$, where the Y_i's are independent and given by

$$Y_0 := \frac{n_x - 1 - \mathrm{E}N(t,x)}{\sigma(N(t,x))}, \quad Y_i := \frac{N(t-\tau, y_i)}{\sigma(N(t,x))}, \quad 1 \le i \le n_x. \tag{6}$$

Lemma 2 *(Liapounov) For any given $\epsilon > 0$, there exists a $\delta = \delta_\epsilon$ such that if $Y_0, \ldots, Y_n$ are random variables satisfying*

1. $|\sum_{i=0}^{n} \mathrm{E}Y_i| \le \delta$,
2. $|\sum_{i=0}^{n} \sigma^2(Y_i) - 1| \le \delta$,
3. $|Y_i - \mathrm{E}Y_i| \le \delta$, for $0 \le i \le n$,

then the distribution function of $\sum_{i=0}^{n} Y_i$ approximates uniformly to within ϵ the normal distribution with zero mean and unit variance.

Recall that

$$\mu(t) = \lim_{x\to\infty} \frac{\sigma^2(N(t,x))}{x}.$$

Lemma 3 *Let $y_1, \ldots, y_n > 0$ and $y_1 + \cdots + y_n = x - (n-1)$. For given $T > 0$, we have that*

$$\lim_{x\to\infty} \frac{1}{x} \sum_{i=1}^{n} \sigma^2(N(t,y_i)) = \mu(t)$$

uniformly in $\mathbf{y} = (y_1, \ldots, y_n)$ and $0 \le t \le T$.

Proof: By (1), we have that, for any $\epsilon > 0$, there exists an $A = A(\epsilon)$ such that

$$\left| \frac{\sigma^2(N(t,x))}{x} - \mu(t) \right| < \epsilon$$

when $x \ge A$ and $0 \le t \le T$. With this in mind, we write

$$\begin{aligned}
&\frac{1}{x}\sum_{i=1}^{n} \sigma^2(N(t,y_i)) - \mu(t) = \\
&\frac{1}{x}\sum_{\{i:y_i \ge A\}} \left[\frac{\sigma^2(N(t,y_i))}{y_i} - \mu(t) \right] y_i + \frac{1}{x}\sum_{\{i:y_i < A\}} \sigma^2(N(t,y_i)) - \frac{\mu(t)}{x}\left[\sum_{\{i:y_i \le A\}} y_i + n - 1 \right] \\
&:= \Sigma_1 + \Sigma_2 + \Sigma_3.
\end{aligned} \tag{7}$$

Now $|\Sigma_1| \le \epsilon$ by our choice of A, and for $y_i < A$, we have $N(t,y_i) \le y_i$, so $\sigma^2(N(t,y_i)) \le A^2$, and hence, $|\Sigma_2| \le nA^2/x$. Finally, $|\Sigma_3| \le B(A+1)n/x$, where $B := \sup_{0\le t\le T} \mu(t)$. From these estimates and (7), we conclude that

$$\left| \frac{1}{x}\sum_{i=1}^{n} \sigma^2(N(t,y_i)) - \mu(t) \right| \le \epsilon + \frac{(A^2 + AB + B)n}{x}, \text{ for all } \mathbf{y} \text{ and } 0 \le t \le T,$$

whereupon the limit $x \to \infty$ yields the lemma. ∎

Lemma 4 *The time by which $n_x - 1$ intervals have been packed in $[0,x]$ satisfies*

$$\tau_x \xrightarrow{d} 0, \text{ as } x \to \infty.$$

Proof: By (1), we know that, for any $\epsilon > 0$ we have $n_x < \mathrm{E}N(\epsilon, x)/2$ with high probability for x large. Then

$$\mathbf{P}(\tau_x > \epsilon) = \mathbf{P}(N(\epsilon, x) < n_x - 1) \leq \mathbf{P}(N(\epsilon, x) < \mathrm{E}N(\epsilon, x)/2)$$

for large x, and by Chebysheff's inequality

$$\mathbf{P}\left(N(\epsilon, x) < \frac{\mathrm{E}N(\epsilon, x)}{2}\right) \leq \frac{4\sigma^2(N(\epsilon, x))}{[\mathrm{E}N(\epsilon, x)]^2},$$

so the lemma follows from the estimates in (1) and (3). ∎

The distribution of $\mathbf{y}$ is identical to its distribution under the packing process of the car parking problem. Hence, as proved in [2], we have

Lemma 5

$$\frac{1}{x} \max_{1 \leq i \leq n} y_i \xrightarrow{d} 0, \text{ as } x \to \infty.$$

We now prove a convergence of moments that leaves us within a short step of the desired CLT.

Theorem 6 *As $x \to \infty$, we have*

$$\sigma^2(Z(t, x)|\mathbf{y}, \tau) \xrightarrow{d} 1, \tag{8}$$

$$\mathrm{E}Z(t, x)|\mathbf{y}, \tau \xrightarrow{d} 0. \tag{9}$$

Proof of (8): We remarked earlier that, for all $\mathbf{y}$ and $\tau \leq t$, the random variable $Z(t, x)|\mathbf{y}, \tau$ is equal in distribution to $\sum_{1 \leq i \leq n} Y_i$. Since Y_0 is constant, we thus have

$$\sigma^2(Z(t, x)|\mathbf{y}, \tau) = \sum_{i=1}^{n} \sigma^2(Y_i) = \frac{\sum_{i=1}^{n} \sigma^2(N(t - \tau, y_i))}{\sigma^2(N(t, x))}. \tag{10}$$

Dividing both numerator and denominator of the last term by x, we obtain from (10), Lemma 3, and (3)

$$\lim_{x \to \infty} \sigma^2(Z(t, x)|\mathbf{y}, \tau) = \frac{\mu(t - \tau)}{\mu(t)} \text{ uniformly in } \mathbf{y} \text{ and } 0 \leq \tau \leq t. \tag{11}$$

Now let $\epsilon > 0$. It is easily verified that $\mu(t)$ is continuous for all t (see e.g. [1]), so we may choose δ_ϵ, $0 < \delta_\epsilon < t$ such that

$$\left|\frac{\mu(t - \tau)}{\mu(t)} - 1\right| \leq \frac{\epsilon}{2} \text{ for } 0 \leq \tau \leq \delta_\epsilon.$$

From this inequality together with (11), we conclude that there exists an $x_\epsilon > 0$ such that

$$|\sigma^2(Z(t, x)|\mathbf{y}, \tau) - 1| \leq \epsilon, \text{ for } x > x_\epsilon, \text{ and } 0 \leq \tau \leq \delta_\epsilon.$$

Hence, for $x > x_\epsilon$, the event $|\sigma^2(Z(t, x)|\mathbf{y}, \tau) - 1| \leq \epsilon$ implies the event $\tau > \delta_\epsilon$ and

$$\mathbf{P}(|\sigma^2(Z(t, x)|\mathbf{y}, \tau) - 1| \leq \epsilon) \leq \mathbf{P}(\tau > \delta_\epsilon).$$

The limit $x \to \infty$ and Lemma 4 then proves (8).

Proof of (9): We have

$$1 = \sigma^2(Z(t,x)) = \mathrm{E}\sigma^2(Z(t,x)|\mathbf{y},\tau) + \mathrm{E}\mathrm{E}^2(Z(t,x)|\mathbf{y},\tau). \tag{12}$$

By (8), we have

$$\liminf_{x\to\infty} \mathrm{E}\sigma^2(Z(t,x)|\mathbf{y},\tau) \geq 1$$

and by (12) we have

$$\limsup_{x\to\infty} \mathrm{E}\sigma^2(Z(t,x)|\mathbf{y},\tau) \leq 1,$$

and so

$$\lim_{x\to\infty} \mathrm{E}\sigma^2(Z(t,x)|\mathbf{y},\tau) = 1$$

Together with (12), this implies

$$\lim_{x\to\infty} \mathrm{E}\mathrm{E}^2(Z(t,x)|\mathbf{y},\tau) = 0$$

which in turn implies (9). ∎

We are now ready for the proof of the CLT.

Proof of Theorem 1: For given $\epsilon > 0$, choose $\delta = \delta_\epsilon$ as in Lemma 2. Define S_x as the intersection of the four events

$$|\sigma^2(Z(t,x)|\mathbf{y},\tau) - 1| \leq \delta, \qquad |\mathrm{E}Z(t,x)|\mathbf{y},\tau| \leq \delta$$
$$\max_{1\leq i\leq n_x} \frac{y_i}{\sigma(Z(t,x)} \leq \delta, \qquad \tau_x \leq t,$$

and let $\overline{S}_x$ denote the complement of S_x. Letting $\Phi(u)$, $-\infty < u < \infty$ denote the standard normal distribution, write

$$\begin{aligned} \mathbf{P}(Z(t,x) \leq u) - \Phi(u) &= \mathbf{P}([Z(t,x) \leq u] \cap S_x) + \mathbf{P}([Z(t,x) \leq u] \cap \overline{S}_x) \qquad (13) \\ &= \left[\int_{S_x} \mathbf{P}(Z(t,x)|\mathbf{y},\tau) - \Phi(u)\right] - \Phi(u)\mathbf{P}(\overline{S}_x + \mathbf{P}([Z(t,x) \leq u] \cap \overline{S}_x). \end{aligned}$$

On S_x, we have $Z(t,x)|\mathbf{y},\tau = \sum_{1\leq i\leq n_x} Y_i$, with the Y_i's independent and satisfying the conditions of Lemma 2. We conclude from (13) and Lemma 2 that

$$\sup_u |\mathbf{P}(Z(t,x) \leq u) - \Phi(u)| \leq \epsilon + 2\mathbf{P}(\overline{S}_x). \tag{14}$$

By Lemmas 2 and 5 and by (1) and (3), we have $\lim_{x\to\infty} \mathbf{P}(\overline{S}_x) = 0$. Theorem 1 thus follows from (14) by first letting $x \to \infty$ and then letting $\epsilon \to 0$. ∎

References

[1] E. G. Coffman, Jr., Leopold Flatto, Predrag Jelenković, and Bjorn Poonen. Packing random intervals on-line (submitted for publication).

[2] A. Dvoretzky and H. Robbins. On the "parking" problem. *MTA Mat. Kut. Int. Kžl.*, 9:209–225, 1964.

[3] J. K. Mackenzie. Sequential filling of a line by intervals placed at random and its application to linear adsorption. *J. Chem. Phys.*, 37(4):723-728, 1962.

[4] A. Rényi. On a one-dimensional random space-filling problem. *MTA Mat. Kut. Int. Kžl.*, 3:109–127, 195 8.