Fluid Approximation to Controlled Jump Markov Processes with Local Transitions

ALEXEY PIUNOVSKIY
The University of Liverpool
Department of Mathematical Sciences
Peach St. Liverpool
L69 7ZL UK
piunov@liverpool.ac.uk

Abstract: Under the heavy traffic condition, i.e. under large values of the scaling parameter \( n \), it is sensible to replace the stochastic queueing model with the corresponding fluid model. After solving an optimal control problem for the latter one, the natural question arises: if we apply the obtained quasi-optimal control strategy to the underlying stochastic system, how far from the optimality will the objective be? We suggest a general method which makes it possible to estimate that accuracy only in the terms of initial data and show that the error approaches zero when the scaling parameter goes to infinity.

Key–Words: Birth-and-death process, Continuous time Markov chain, Fluid model, Optimal control, Queuing system, Dynamic programming, \( \mu C \)-rule, Inventory

1 Introduction

Stochastic jump processes, especially birth-and-death processes, are widely used in the queuing theory, computer networks and information transmission. The state of such process describes the instant length of the queues (numbers of packets at different edges to be transmitted through the net). Under the heavy traffic conditions, trajectories of such processes are close to the trajectories of deterministic dynamic systems. Therefore, if we consider the related optimal control problems, we expect that the optimal control strategy in the deterministic (‘fluid’) model will be nearly optimal in the underlying stochastic model.

Justification of the fluid approximation in different settings can be found in [6, 8]. On the other hand, very often the fluid model is used, without justification, for the study of real life processes that are of the birth-and-death type [1, 2]. In the current report, a new technique for calculating the accuracy of this approximation is described. In a nutshell, instead of the study of trajectories, we investigate the corresponding dynamic programming equations. It should be emphasized that we deal also with multiple-dimensional lattices, so that the results are applicable to complex communicating systems of queues. Other areas of application are population dynamics, mathematical epidemiology, and inventory systems.

In the next section, we present very briefly the results for a pure birth-and-death process. A random walk on a multiple-dimensional lattice is considered in Section 3, in the framework of dynamic priorities allocation to different types of jobs served by a single server. The proofs, omitted due to the space limitation, are based on the dynamic programming approach and the Dynkin’s formula. Application to inventory theory is presented in Section 4. Actually, only Sections 2 and 3 fit the optimal control theory, in Section 4 we deal with the parametric optimization.

2 One-dimensional lattice

Following the standard practice, the controlled birth-and-death process \( Y_t \) is defined by the following elements:

- \( S = \{0, 1, \ldots \} \) is the state space,
- \( A \) is the action space (arbitrary Borel),
- \( Q = [q_{i,j}(a)] \) is the tri-diagonal matrix of transition rates; \( q_{i,j} = 0 \) if \( |i-j| > 1 \);
- \( G(i, a) \) is the (real, measurable) loss rate; \( G(0, a) \equiv 0 \).

A control strategy \( \Phi \) is a mapping from \( S \) to \( A \). We restrict ourselves to stationary nonrandomised
strategies because, under rather general conditions, they are sufficient for solving optimization problems.

The optimal control problem under consideration looks like follows:

\[ W^\Phi(i) = E_i^\Phi \left( \int_0^\infty G(Y_s, \Phi(Y_s))ds \right) \rightarrow \inf_{\Phi} \cdot \quad (1) \]

Here \( E_i^\Phi \) is the expectation on the space of trajectories \( \{Y_s\}_{s\geq0} \) starting from \( Y_0 = i \) and absorbing at zero, w.r.t probability measure generated by strategy \( \Phi \). The rigorous mathematical constructions can be found in [5].

In fact, we shall consider a sequence of described models, so that all their parameters \( n\Lambda, nM \) and \( nG \), as well as objective \( nW^\Phi \) will be indexed with \( n = 1, 2, \ldots \) Namely, we assume that measurable functions \( \lambda(y, a), \mu(y, a), \) and \( g(y, a) \) are fixed for \( y > 0, a \in \Lambda \), such that \( \mu(y, a) > \lambda(y, a); \lambda(0,a) = \mu(0,a) = g(0,a) \equiv 0; \) and

\[ n\Lambda_i(a) = n\lambda(i/n, a), \quad nM_i(a) = n\mu(i/n, a), \quad (2) \]

\[ nG_i(a) = g(i/n, a). \]

This is the standard fluid scaling: see [6, 8]. Now one can introduce the (fluid) absorbing optimal control problem

\[ \frac{dy}{d\tau} = \lambda(y, a) - \mu(y, a); \quad \int_0^\infty g(y, a)d\tau \rightarrow \inf_{a(\cdot)} \]

where the infimum is taken over all control strategies \( a(\cdot) \).

Roughly speaking, if \( a = \varphi^*(y) \) is an optimal feedback control strategy in problem (3) then, for stationary strategy \( \Phi^*(i) = \varphi^*(i/n) \), inequality

\[ \sup_{0 \leq \varepsilon \leq \gamma_m} \left| nW^{\Phi^*}(i) - \inf_{\Phi} nW^\Phi(i) \right| \leq \varepsilon(n) \]

holds, where, for any value of \( \gamma \), one can provide an explicit formula for \( \varepsilon(n) \), in terms of the initial data, and \( \varepsilon(n) = O(1/n) \rightarrow 0 \) as \( n \rightarrow \infty \). The detailed proofs and the exact formula for \( \varepsilon(n) \) can be found in [9] and in the forthcoming article [10].

Example. Consider the M/M/1 queueing system with the controlled input stream \( n\lambda(j/n, a) = n(d_0 + d_1 a), d_0, d_1 > 0, d_0 + d_1 < 1, a \in \Lambda = [0, 1]; \) the service intensity \( n\mu(j/n, a) = n \) is constant. As usual, \( n \) is a fixed large enough parameter. The initial state is \( i > 0 \), and we observe the trajectory up to the absorption at zero. One can consider this model as a version of call admission control. The server always accepts the jobs from one stream with intensity \( nd_0 \), but can chose any probability \( a \) of accepting jobs from another stream, intensity \( nd_1 \). Suppose we are interested in the total expected throughput (to be maximised), as well as the total expected queue length (to be minimised). Therefore, \( nG(i, a) = i/n - Ra \), where \( R > 0 \) is a given constant (Lagrange multiplier), and one has to solve problem (1). The corresponding fluid model is defined by

\[ \lambda(y,a) = d_0 + d_1 a, \quad \mu(y,a) = 1, \quad g(y,a) = y - Ra. \]

The Bellman function for it looks as follows:

\[ v(y) = \begin{cases} Ry - y^2/2 - d_0 - d_1 & \text{if } 0 \leq y \leq y^*; \\ y^2 - (y^*)^2/2 - d_0 - d_1 & \text{if } y > y^*, \\ 0 & \text{if } y = y^*, \\ \end{cases} \]

Feedback control strategy

\[ \varphi^*(y) \equiv \begin{cases} 1, & \text{if } 0 < y \leq y^*; \\ 0, & \text{if } y > y^*; \\ \end{cases} \]

is optimal for the problem (3)

Let us fix \( d_0 = 0.25; d_1 = 0.5; R = 1 \). Now, after choosing \( \gamma = 5 \), the calculations show that \( \varepsilon(n) = 0.022 \) for \( n = 100,000 \). It should be emphasized that this estimate of the accuracy of the fluid approximation is very rough.

### 3 Multiple-dimensional lattice: \( \mu C \)-rule

Suppose there are \( m \) types of jobs to be served by a single server. Arrival and service rates for type \( j \) equal \( n\lambda_j \) and \( n\mu_j \), where, like previously, \( n \) is the scaling parameter. We shall consider the Markovian case when all the service and inter-arrival times are exponential; we assume also that there is infinite space for waiting. The holding cost of one type \( j \) job equals \( C_j/n \) per time unit. At any moment, the server should chose a job for service from the queue: that is, the action \( a = j \in \{1, 2, \ldots, m\} \) means that a type \( j \) job is under service. Note that we allow to switch the server to another job even before the service is completed. The goal is to minimize the total holding cost up to the absorption at the zero state (empty queue).

Mathematical problem looks as follows.

\[ S = \{(i_1, i_2, \ldots, i_m)\} \] is the state space; \( i_j \geq 0 \) equals the number of jobs of type \( j \) in the system.
where $A = \{1, 2, \ldots, m\}$ is the action space.

If $Y = (Y_1, Y_2, \ldots, Y_m)$ is the current state, then only the following transitions can occur:
- one job of type $j$ arrives; transition rate to the new state $Y' = (Y_1, Y_2, \ldots, Y_j + 1, \ldots, Y_m)$ equals $n\lambda_j$;
- if $Y_j > 0$ and $a = j$ then the service can be completed; transition rate to the new state $Y' = (Y_1, Y_2, \ldots, Y_j - 1, \ldots, Y_m)$ equals $n\mu_j$.

State $Y = 0$ is absorbing.

In what follows, we accept that $a \neq j$ in case $Y_j = 0$: there is no reason to serve a dummy job.

Finally, $G(Y, a) = \frac{1}{n} \sum_{j=1}^{m} C_j Y_j$, and we study problem (1). Of course, initial state $(i_1, i_2, \ldots, i_m)$ and the process $Y_s$ are now multiple-dimensional; the objective to be minimized is denoted as $nW^\Phi(i_1, i_2, \ldots, i_m)$.

**Conditions 1** $\sum_{j=1}^{m} \frac{\lambda_j}{\mu_j} < 1$, i.e. this queueing system is stable.

The corresponding fluid model is described by the following equations
\[ \frac{dy}{d\tau} = f(y, a), \]
where $y \in \mathbb{R}^m_+$, $a \in A$, and, for $y \neq 0$,
\[ f_j(y, a) = \begin{cases} \lambda_j, & \text{if } j \neq a, \text{ or if } y_j = 0, \\ \lambda_j - \mu_j, & \text{if } j = a \text{ and } y_j > 0. \end{cases} \]

If $y = 0$ then $f(y, a) = 0$.

Performance functional:
\[ F = \int_0^\infty g(y)d\tau \to \inf_{a(\cdot)} \] (5)
where $g(y) = \sum_{j=1}^{m} C_j y_j$.

Without loss of generality, further we assume that
\[ \mu_1 C_1 \geq \mu_2 C_2 \geq \ldots \geq \mu_m C_m \]
and introduce function
\[ v(y_1, y_2, \ldots, y_m) \triangleq v_m(y_1, y_2, \ldots, y_m) \]
defined by the (recursive) formulae
\[ v_1(y_1) = \frac{C_1 y_1^2}{2(\mu_1 - \lambda_1)}; \]
\[ T_1(y_1) = \frac{y_1}{\mu_1 - \lambda_1}; \]
\[ v_{k+1}(y_1, y_2, \ldots, y_{k+1}) = v_k(y_1, y_2, \ldots, y_k) + C_{k+1} \left[ y_{k+1} T_k(y_1, y_2, \ldots, y_k) \right. \]
\[ \left. + \frac{\lambda_{k+1} T_k^2(y_1, y_2, \ldots, y_k)}{2} \right) \]
\[ + \frac{\lambda_{k+1} T_k(y_1, y_2, \ldots, y_k)}{2(\eta_{k+1} - \lambda_{k+1})} \]
\[ T_{k+1}(y_1, y_2, \ldots, y_{k+1}) = T_k(y_1, y_2, \ldots, y_k) \]
\[ + \frac{\lambda_{k+1} T_k(y_1, y_2, \ldots, y_k)}{\eta_{k+1} - \lambda_{k+1}}. \]

Here $\eta_{k+1} = \mu_{k+1} \left( 1 - \frac{\lambda_k}{\mu_k} - \ldots - \frac{\lambda_2}{\mu_2} \right)$ is the effective service rate of $(k+1)$-type jobs under the feedback $\mu C$-strategy
\[ a = \varphi^*(y) \triangleq (k + 1) \]
\[ \times I\{y_1 = 0, y_2 = 0, \ldots, y_k = 0, y_{k+1} > 0\}, \]
where $I$ stands for the indicator function. This strategy allows to serve type $(k+1)$-jobs only if there are no jobs of types $1, 2, \ldots, k$. Function $v_k$ coincides with the Bellman function in case there are only the first $k$ types of jobs, $T_k$ is the corresponding time until absorption at zero.

**Lemma 1** Under Condition 1, the $\mu C$-strategy $\varphi^*$ is optimal in the fluid model (4), (5).

Note that, under the control strategy $\varphi^*$, if $y_1 = y_2 = \ldots = y_k = 0$ and $y_{k+1} > 0$ then the first $k$ components remain zero, but the dynamics on that hyperplane is in the sliding mode which can be equivalently described by equations
\[ \frac{dy_j}{d\tau} = 0, \ j \leq k; \quad \frac{dy_{k+1}}{d\tau} = \lambda_{k+1} - \eta_{k+1}; \]
\[ \frac{dy_j}{d\tau} = \lambda_j, \ j > k + 1. \]

**Theorem 1** For the feedback $\mu C$-strategy
\[ \Phi^*(Y) \triangleq \varphi^*(Y/n), \]
for any vector $\hat{y} \in \mathbb{R}^m_+$, the following inequality holds:
\[ \sup_{0 \leq (i_1, i_2, \ldots, i_m) \leq \hat{y}} |nW^\Phi(i_1, i_2, \ldots, i_m) - v(i_1/n, i_2/n, \ldots, i_m/n)| \]
\[ \leq \frac{mD}{n} \left( \max_{1 \leq j \leq m} \lambda_j \right) \sum_{j=1}^{m} \eta_j - \lambda_j \prod_{j=1}^{m} \eta_j - \lambda_j, \]
where $D = \max_{1 \leq j \leq m} \frac{\partial^2 v}{\partial y_j^2}$ is a constant since function $v$ is quadratic; vector inequalities are component-wise.
For all large enough $n$, strategy $\Phi^*$ is nearly optimal for all initial states $(i_1, i_2, \ldots, i_m) \leq \tilde{y}n$ in the stochastic problem (1). Namely,

\[
\sup_{0 \leq (i_1, i_2, ..., i_m) \leq \tilde{y}n} nW^{\Phi^*}(i_1, i_2, ..., i_m) - \inf_{\Phi} nW^\Phi(i_1, i_2, ..., i_m) \leq \frac{2mD}{n} \left( \max_{1 \leq j \leq m} \lambda_j \right) \sum_{j=1}^{m} \frac{\tilde{y}_j}{\eta_j} - \frac{1}{\eta_l} \prod_{l=j+1}^{m} \eta_l - \lambda_j
\]

(8)

It is known that the $\mu C$-strategy (6) is optimal for the stochastic discounted problem. The asymptotic optimality of the $\mu C$-strategy under the ‘heavy traffic’ conditions (very close to our setting with a large value of $n$) was established in [7]. At the same time, rather often the $\mu C$-strategy is counter-intuitively not optimal. Let us consider, for example, the absorbing stochastic model investigated above, with $m = 2$, and under the finite buffer condition: there cannot be more than one job of each type in the system. Assume for simplicity that $\mu_1 = \mu_2 = \mu$. One can prove that the $\mu C$-strategy is optimal iff $C_1 - C_2 \geq \left( C_1 + C_2 \right) \frac{\lambda_2 - \lambda_1}{\mu}$. In case $\lambda_1 \gg \lambda_2$, it is worth serving the second type job (even if $C_1 \mu > C_2 \mu$) because after that, while serving the first type, it is unlikely that a job of type 2 will arrive, and we can quickly reach the absorbing zero state.

Let us have a quick look at the following long-run average modification of the stochastic model. After reaching the zero state, the $Y_i$ process can jump to states $(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)$ with probabilities $\lambda_1 / \left( \sum_{j=1}^{m} \lambda_j \right), \lambda_2 / \left( \sum_{j=1}^{m} \lambda_j \right), \ldots, \lambda_m / \left( \sum_{j=1}^{m} \lambda_j \right)$. Therefore, under a stationary control strategy $\Phi$, the expected total loss, between the two consecutive visits to the zero state, equals

\[
\frac{1}{\sum_{j=1}^{m} \lambda_j} \left[ \lambda_1 W^\Phi(1, 0, \ldots, 0) + \lambda_2 W^\Phi(0, 1, 0, \ldots, 0) + \ldots + \lambda_m W^\Phi(0, 0, \ldots, 1) \right]
\]

(9)

Here and below we omit the $n$ index because it is fixed. Denoting $U^\Phi(i_1, i_2, \ldots, i_m)$ the expected time to hit zero starting from state $(i_1, i_2, \ldots, i_m) \neq 0$, we see that the expected time interval, between the two consecutive visits to zero, equals

\[
\frac{1}{\sum_{j=1}^{m} \lambda_j} \left[ \lambda_1 U^\Phi(1, 0, \ldots, 0) + \lambda_2 U^\Phi(0, 1, 0, \ldots, 0) + \ldots + \lambda_m U^\Phi(0, 0, \ldots, 1) \right]
\]

(10)

We aim to minimize the ratio of (9) to (10) which is the same as the long-run average loss

\[
\lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left[ \int_0^T G(Y_t, \Phi(Y_t)) dt \right],
\]

because process $Y_t$ is regenerative [11].

Let us show that the $\mu C$-strategy is not always optimal in this problem. Indeed, suppose $\mu_1 C_1 = \mu_2 C_2$, and consider the two $\mu C$-strategies: $\Phi_1$ giving the priority to type 1 and $\Phi_2$ giving the priority to type 2. Suppose expression (10) is smaller at $\Phi_1$ and $\Phi_2$, so that strategy $\Phi_2$ is better than $\Phi_1$. One can prove that the ratio of (9) to (10) is less at $\Phi_1$, which is identical to the value of (9) at $\Phi_2$, will become only slightly smaller: function (9) is continuous wrt $C_1$. In this case, strategy $\Phi_2$ remains better than $\Phi_1$, the latter being the proper $\mu C$-strategy.

Controlled fluid models of general communication networks with linear holding cost $\sum_{j=1}^{m} C_j Y_j$ were studied in [2]. Optimality of the $\mu C$-strategy established in the above Lemma 1 is consistent with the more general result [2]: there exists a finite collection of polyhedral cones, covering the total state space $\mathbb{R}_+^m$, such that the value of the optimal feedback control strategy is constant inside each of those cones. At the same time, the optimal fluid strategy, translated back to the underlying stochastic network using formula (7), can be far not optimal: see [3], where a simple example of a tandem queue was discussed. In such queues, the $\mu C$-strategy can also be not optimal, as it was shown in [4].

4 Applications to the inventory theory

We start with a rather general fluid model of an inventory system, which is a particular case of the model studied in [1]. If the inventory level is $y \geq 0$ then the demand rate is $\mu(y) > 0$, so that

\[
\frac{dy}{dt} = -\mu(y).
\]

At the moment $\tau^*$ when $y(\tau^*)$ reaches zero, the cycle is over and $y(\tau^* + 0) = z > 0$, i.e. the replenishment is instantaneous, the set-up cost being $K$. Holding $y$ units results in the cost $g(y)$ per time unit. If we take into account also the profit, then one has to adjust function $g$ by subtracting $c(y)\mu(y)$, where $c(y)$ is the profit from selling one unit. We are interested in minimizing the (long-run) total cost per unit time (tcu):

\[
tcu(z) = \lim_{T \to \infty} \frac{1}{T} \left\{ \int_0^T g(y(\tau)) d\tau + K \left[ T \frac{T}{T_z} \right] \right\}
\]
Like previously, we shall estimate the difference \( C \) with Lipschitz constants \( t \) in (small) units, so that the state space \( S \) is the solution to equation
\[
\int_0^{T_c} \mu(y(\tau)) d\tau = z, \quad \text{i.e.} \quad T_c = \int_0^z \frac{dy}{\mu(y)}.
\]
The best possible value \( z^* \) is called economic order quantity (eoq).

All the introduced integrals are well defined under the following conditions.

**Conditions 2**

(a) There exist constants \( C_1, C_2, \) and \( \delta \) such that
\[
\delta \leq \mu(y) \leq C_1, \quad |g(y)| \leq C_2.
\]
(b) Functions \( g(y) \) and \( \mu(y) \) are Lipschitz continuous with Lipschitz constants \( C_3 \) and \( C_4 \) correspondingly.

The corresponding stochastic model looks like follows. The product is measured and demanded in (small) units, so that the state space \( S = \{0, 1, 2, \ldots, Z = \lfloor nz \rfloor \} \); the square brackets stay for the integer part; \( n \) is the (big) scaling parameter. The random process \( Y_s \) under consideration represents the number of units of product in stock at time moment \( s \). Transition rates equal \( q_{i,i-1} = n\mu(i/n), \) if \( i \geq 1, \) and \( q_{0,Z} = n\lambda; \) the lead time is exponential with parameter \( n\lambda. \) All the remainder values \( q_{i,j \neq i} = 0. \) The loss rate is
\[
^nG(i) = g(i/n) + Kn\mu(1/n)I \{i = 1\}.
\]
The last term corresponds to the expected set-up cost at the moment when \( Y_n = 0. \) Performance functional:
\[
^nTCU(Z) = \lim_{T \to \infty} \frac{1}{T} E \left[ \int_0^T \!^nG(Y_s) ds \right] \to \inf \begin{array}{ll}
\int_{Z > 0} & . \end{array}
\]
Like previously, we shall estimate the difference \( \lfloor nTCU([nz]) - tcu(z) \rfloor \) and prove that \( \text{eoq} z^* \) provides a nearly optimal value \( [nz]* \) to \( Z. \)

**Theorem 2** Suppose the order size \( z > 0 \) is fixed and Conditions 2 hold for \( y \in [0, z]. \) Then
\[
\left| \lfloor nTCU([nz]) - tcu(z) \rfloor \right| \leq \frac{C_4^2}{n(z - 1/n)} \delta^4.
\]
Under the conditions of the following Corollary, the values \( z < \frac{\delta^2 K}{C_1^2 C_2} \) and \( Z < \frac{\delta^2 K}{C_1 C_2} - \frac{\delta}{n\lambda} \) cannot be optimal in the fluid and stochastic models correspondingly.

**Corollary 1** If \( g(z) > 0 \) and \( N \) is such a number that \( \frac{\delta^2 K}{C_1^2 C_2} - \frac{\delta}{n\lambda} > 0, \) then, for all \( n \geq N, \) for all \( z \geq \frac{\delta^2 K}{C_1 C_2} - \frac{\delta}{n\lambda}, \) the following inequality holds
\[
\left| \lfloor nTCU([nz]) - tcu(z) \rfloor \right| \leq \frac{E}{n}
\]
where
\[
E = \frac{C_3^2 C_2 N\lambda}{(\delta^2 K) N\lambda - \delta C_1 C_2 - C_2 C_2) \delta^4}
\]
\[
\times \left\{ \left( C_2 + \frac{K\delta C_1 C_2 N\lambda}{\delta^2 K N\lambda - \delta C_1 C_2} \right) \right.
\]
\[
\times \left( C_4 \cdot \frac{\delta^2 K N\lambda - \delta C_1 C_2}{C_1 C_2 N\lambda} + \delta + \frac{\delta^2}{\lambda} \right)
\]
\[
+ \frac{\delta^2 K N\lambda - \delta C_1 C_2}{C_1 C_2 N\lambda} \left( C_2 C_4 + C_1 C_3 \right) + C_2 \delta + \frac{\delta^2}{\lambda} \right\}
\]
We see that, if we find the \( \text{eoq} z^* \) then, for \( n \geq N, \)
\[
\left| \lfloor nTCU([nz]) - \inf_{z} nTCU(Z) \right| \leq \frac{2E}{n}
\]
Note that it is usually much easier to find \( z^* \) than the \( \text{EOQ} Z^* \) for the stochastic model, and if we accept the value of \( [nz]* \) then, for big \( n, \) \( nTCU([nz]) \) will be close to the best possible.

In the classical case, when \( \mu(y) = D \) and \( g(y) = hy, \) we obtain:
\[
tcu(z) = \frac{KD}{z} + \frac{hz}{2}
\]
and
\[
^nTCU(Z) = \frac{\lambda(hz^2 + hz + 2n^2 DK)}{2(\lambda nZ + nD)}.
\]
Under fixed $n$ and $Z$, formulae

\[
P(0) = \frac{1}{n\lambda} / \left( \sum_{j=1}^{Z} \frac{1}{n\mu(j/n)} + 1 \right),
\]
\[
P(i) = \frac{1}{n\mu(i/n)} / \left( \sum_{j=1}^{Z} \frac{1}{n\mu(j/n)} + 1 \right),
\]
\[
i = 1, 2, \ldots, Z
\]
provide the stationary distribution for the jump random process $Y_s$.

At the same time, in the fluid model the limiting invariant probability density is given by expression

\[
p(y) = \frac{1}{\mu(y)} / \int_{0}^{y} \frac{du}{\mu(u)}
\]

If $Z = [n\lambda]$, then

\[
|P(0)| \leq C_1 \frac{\delta}{n(z - 1/n)\lambda}
\]

and, for $i = 1, 2, \ldots, [n\lambda]$,

\[
\left| P(0) - \int_{(i-1)/n}^{i/n} p(u)du \right| \leq C_1 \frac{2\zeta C_4 + \delta + \theta^2}{z(z - 1/n)n^2\delta^3}.
\]

Therefore, again, if $n$ is large, one can very precisely estimate the distribution $P$ based on the density $p$ for the fluid model.

5 Conclusion

The convergence of trajectories of jump processes with local transitions to those of the corresponding ‘fluid’ dynamic systems was previously proved based on the Law of Large Numbers (see e.g. [6]). In the present work, we provide the rate of that convergence in terms of the objective functionals, in the framework of controlled models, and present the explicit formulae for the error term, based only on the initial data. Meaningful examples show that the theory developed can be applied to many real life situations. Another field of applications is population dynamics and mathematical epidemiology, which is not touched in the current report. All the proofs were omitted here due to the space limitation. They were based on the dynamic programming and the Dynkin’s formula, and will be published in the nearest future.

Acknowledgements: The author is thankful to the Royal Academy of Engineering for supporting the travel to Tbilisi.

References:


