Simple proof of Dynkin’s formula for single–server systems and polynomial convergence rates

A. Yu. Veretennikov†  G. A. Zverkina‡

May 7, 2014

Abstract

An elementary rigorous justification of Dynkin’s identity with an extended generator based on the idea of a complete probability formula is given for queueing systems with a single server and discontinuous intensities of arrivals and serving. This formula is applied to the analysis of ergodicity and, in particular, to polynomial bounds of convergence rate to stationary distribution.

1 Introduction

In the last decade, systems generalising $M/G/1/\infty$, or simply $M/G/1$ (cf. [8]) – one of the oldest and most popular queueing systems – attracted much attention, see [1] – [7], [16]. In this paper a single–server system similar to [18] is considered, in which intensities of new arrivals as well as of their serving could depend on the “whole state” of the system and the whole state includes the number of customers in the system – waiting and on service – and the elapsed time of the last serving. Batch arrivals are not allowed. Emphasize that the elapsed serving time is assumed known at any moment, but not remaining service times for each customer. For definiteness, the discipline of serving may be regarded as FIFO, although, this is of no importance for the questions addressed in this paper: all results are valid for any serving discipline without time sharing. Recall that despite a non-markovian character of the model, this model – as any other process – may be made Markov by extending the state...

---

*School of Mathematics, University of Leeds, LS2 9JT, Leeds, UK, & Institute of Information Transmission Problems, B. Karetny 19, 127994, Moscow, Russia, email: a.veretennikov @ leeds.ac.uk
†The first author’s work was partially supported by RFBR grant 13-01-12447 ofi_.m2.
‡Department of Applied Mathematics, Moscow State University of Railway Engineering (MIIT), Obraztsova 9, build. 9, 127994, GSP-4, Moscow, Russia, email: zverkina @ gmail.com
Dynkin’s formula, or Dynkin’s identity [6, Ch. 1, §3],

$$E_X f(X_t) - f(X) = E_X \int_0^t \mathcal{G} f(X_s) \, ds$$  \hspace{1cm} (1)

for $f$ from some appropriate class of functions and its non-homogeneous counterpart

$$E_X \varphi(t, X_t) - \varphi(0, X) = E_X \int_0^t \left( \frac{\partial}{\partial s} \varphi(s, X_s) + \mathcal{G} \varphi(s, X_s) \right) \, ds,$$  \hspace{1cm} (2)

for appropriate functions of two variables ($\varphi(t, X)$) play a very important role in the analysis of Markov models. Here $\mathcal{G}$ is understood as a generator or infinitesimal operator of the process, and $X$ is a (non-random) initial value of the process. Note that the formula (2) may be considered as a version of (1) written for the couple $(t, X_t)$. In this paper it is shown that both formulae (1–2) hold true for a large class of functions $f, \varphi$ with an extended generator $\mathcal{G}$ (see below) and under the minimal assumptions on regularity of intensities (i.e., without any regularity). For short, in the title it was called Dynkin’s identity, but, in fact, Dynkin’s formula with generalised generators will be established. As an application, convergence rate bounds to stationary distribution are established under appropriate recurrence conditions similar to those in [18], but without regularity assumed therein. Minimal or no regularity may be important, in particular, in queueing systems with control, where optimal control is usually discontinuous.

This model may be considered as a partial case of piecewise deterministic Markov processes (PDMP) considered in [5], [8], [9, Ch. 7], [10], et al. Notice that the method in [5,9] is based on martingale theory and requires some further references. In this paper we aim to show that the “naive” approach based on a properly designed version of a complete probability formula works well, at least, in our particular model and may be presented in a self–contained way. Also notice that our model allows a presentation as a piecewise linear Markov process (PLMP) as in [8], [10], while [5] and [9] study a more general deterministic behaviour between “Markov” jumps.

The standard proof of the formulae (1–2) given in [6] in “good cases” is practically trivial; however, this triviality is based on the notion of generator (not extended one), which does assume some regularity. For example, in the theory of diffusion processes a similar notion requires continuous coefficients of drift and diffusion; at the same time, it is well-known that Dynkin’s formula itself still holds true for diffusion processes with a weaker notion of extended generator. This is possible due to PDE theory results, namely, Alexandrov’s inequality and Krylov’s estimate. In queueing theory there are no similar resources and, as a consequence, in some classical works a strict justification of Dynkin’s formulae is just dropped. In the authors’ view, this issue must be addressed and the goal of this paper is to show that such a rigorous justification is well possible practically without – or with a minimum of – martingale
theory. As a first application, convergence rate bounds to the stationary regime is shown for a class of models generalising the system $M/G/1$ similar to [18] but for a wider class of intensities. This may be useful in some other areas such as stochastic averaging and Poisson equations, see, for example, [12]. The bounds in [18] were proposed as a complement to the results from [13], [19], et al. about convergence to stationarity in non-Markov queueing models, in particular, of Erlang–Sevastyanov’s type. The authors hope to apply this approach to such systems as well.

The paper consists of the Section 1 – Introduction, of the setting and main results in the Section 2, of a set of auxiliary results in the Section 3, of the proof of Dynkin’s formulae in the Section 4, of the proof of some extension in the Section 5 and of a brief reminder of the techniques leading to convergence rate bounds in the Section 6.

2 The setting and main results

2.1 Dynkin’s formulae and generalised generator

Recall [6] that the generator of a Markov process $(X_t, t \geq 0)$ is an operator $\mathcal{G}$, such that for a sufficiently large class of functions $f$

$$\sup_X \lim_{t \to 0} \left\| \frac{E_X f(X_t) - f(X)}{t} - \mathcal{G} f(X) \right\| = 0$$

(3)

in the norm of the state space of the process. The notion of generator does depend on this norm. However, in this paper the classical generator is not used, so we skip this point here. Recall further that the class of functions for which (3) holds true is called domain of a generator (of course, in general, it depends on the norm) and in most cases its full description is a hard task. Fortunately, often it suffices to have a wide enough class of such functions. On the other hand, an operator $\mathcal{G}$ is called extended generator if (3) is replaced by its corollary (1), also for a wide enough class of functions $f$. It turns out that in many situations this notion suffices.

Let us present the class of models under investigation in this paper. The state space is a union of subspaces,

$$\mathcal{X} = \{(0,0)\} \bigcup \{(n,x) : n = 1, 2, \ldots, N, x \geq 0\}$$

$(1 \leq N \leq \infty)$ with topology arising from the metric $(X = (n, x), X' = (n', x'))$

$$\text{dist}(X, X') = |n - n'| + |x - x'|.$$

Functions of class $C^1(\mathcal{X})$ are understood as functions with classical continuous derivatives with respect to the variable $x$. Functions with compact support on $\mathcal{X}$ are understood as functions vanishing outside some domain bounded in this metric: for example, $C^1_0(\mathcal{X})$ stands for the class of functions with compact support and one continuous derivative. There is a generalised Poisson arrival flow with intensity $\lambda(X)$,
where \( X = (n, x) \) for \( n > 0 \), and \( X = (0, 0) \) for \( n = 0 \); group arrival are not allowed. If \( n > 0 \), then the server is serving one customer while all others are waiting in a queue; if \( n = 0 \) then the server remains idle until the next arrival and the intensity of such arrival at the state \((0, 0)\) is constant. Here \( n \) denotes the total number of customers in the system and \( x \) stands for the elapsed time of the current serving.

Denote \( n_t = n(X_t) \) – the number of customers corresponding to the state \( X_t \) and \( x_t = x(X_t) \), the second component of the process \((X_t)\), i.e. the elapsed time of the current serving. If \( n = 0 \), then there is no serving. Below, we will investigate formally only the case with infinite number of waiting places, but the situation where this number is bounded may be included by assuming \( \lambda(n, x) \equiv 0 \) for any \( n \) large enough. For any \( X = (n, x) \), intensity of serving \( h(X) \equiv h(n, x) \) is defined; it is also convenient to assume \( h(X) = 0 \) for \( n(X) = 0 \). Both intensities \( \lambda \) and \( h \) are understood in the following way, which is a definition: on any nonrandom interval of time \([t, t+\Delta)\), conditional probability given \( X_t \) that the current serving will not be finished and there will be no new arrivals reads,

\[
\exp \left( -\int_0^\Delta (\lambda + h)(n_t, x_t + s) \, ds \right). \tag{4}
\]

This is similar to the approach developed in [5]; the main difference is that in this paper the basis for the proof of Dynkin’s identity is a complete probability (or complete expectation) formula rather than martingale problem. In the sequel, \( \lambda \) and \( h \) are assumed to be bounded. In this case, for \( \Delta > 0 \) small enough, the expression in (4) may be rewritten as

\[
1 - \int_0^\Delta (\lambda + h)(n_t, x_t + s) \, ds + O(\Delta^2), \quad \Delta \to 0, \tag{5}
\]

and this what is “usually” replaced by

\[
1 - (\lambda(X_t) + h(X_t))\Delta + O(\Delta^2).
\]

In our situation, the latter replacement may be incorrect because of discontinuities of the functions \( \lambda \) and \( h \). Emphasize that from time \( t \) until the next jump, the evolution of the process \( X \) is deterministic. The (conditional given \( X_t \)) density of the moment of a new arrival or of the end of the current serving after \( t \) at \( x_t + z, z \geq 0 \) equals,

\[
(\lambda(n_t, x_t + z) + h(n_t, x_t + z)) \exp \left( -\int_0^z (\lambda + h)(n_t, x_t + s) \, ds \right). \tag{6}
\]

Further, given \( X_t \), the moments of the next “candidates” for jumps up and down are
conditionally independent and have the (conditional – given $X_t$) density, respectively,

$$\lambda(n_t, x_t + z) \exp \left( - \int_0^z \lambda(n_t, x_t + s) \, ds \right)$$

and

$$h(n_t, x_t + z) \exp \left( - \int_0^z h(n_t, x_t + s) \, ds \right), \ z \geq 0.$$  \hspace{1cm} (7)

Notice that (6) does correspond to conditionally independent densities given in (7).

For modelling the evolution, both candidate moments should be realised and the minimal of the two chosen, which will determine whether the next jump will be, indeed, up or down. The component $n_t$ is, generally speaking, not Markov, however, the couple $(X_t = (n_t, x_t))$ in the state space $\mathcal{X}$ is Markov by construction and we stress out that continuity of $\lambda$ or $h$ was not used. The distribution of the process in the space of trajectories is uniquely determined, that is, if any other construction is suggested, yet the distributions of all moments of next jumps up and down ought to be the same.

A natural candidate to the role of the extended generator is (see a short comment in the beginning of the proof of the Theorem 1)

$$\mathcal{G}f(X) := \frac{\partial}{\partial x} f(X) + \lambda(X)(f(X^+) - f(X)) + h(X)(f(X^-) - f(X)),$$

(recall that $h(0,0) = 0$), where for any $X = (n, x), X^+ := (n + 1, x), \ X^- := ((n - 1) \vee 0, 0)$

(here $a \vee b = \max(a, b)$).

**Theorem 1** If the functions $\lambda$ and $h$ are Borel measurable and bounded, then the formulae (7) and (2) hold true for any $t > 0$ for every $f \in C^1_b(\mathcal{X})$ and $\varphi \in C^1_b([0, \infty) \times \mathcal{X})$, respectively. Moreover, the process $(X_t, t \geq 0)$ is strong Markov with respect to the filtration $(\mathcal{F}^X_t, t \geq 0)$.

Let

$$L_m(X) = (n + 1 + x)^m; \quad L_{k,m}(t, X) = (1 + t)^k L_m(X).$$

The following extensions of Dynkin’s formulae for unbounded functions hold true: it will be needed in the proof of the Theorem 2 even though the main body of this proof will be hidden.
Corollary 1 Under the assumptions of the Theorem 1,

\[ L_m(X_t) - L_m(X) = \int_0^t \lambda(X_s) \left[ (L_m(X_s^+) - L_m(X_s)) \right. \]

\[ + h(X_s) \left( L_m(X_s^-) - L_m(X_s) \right) + \frac{\partial}{\partial x} L_m(X_s) \right] ds + M_t, \]

with some martingale \( M_t \), and also

\[ L_{k,m}(t, X_t) - L_{k,m}(0, X) = \int_0^t \left[ \lambda(X_s) (L_{k,m}(s, X_s^+) - L_{k,m}(s, X_s)) \right. \]

\[ + h(X_s) (L_{k,m}(s, X_s^-) - L_{k,m}(s, X_s)) + \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial s} \right) L_{k,m}(s, X_s) \right] ds + \tilde{M}_t, \]

with some martingale \( \tilde{M}_t \).

About martingale language in queueing models see, for example, [9] or [13].

2.2 Stability and convergence rate

Denote

\[ \Lambda := \sup_{n, x: n>0} \lambda(n, x). \]

Recall that the process has no explosion with probability one due to the boundedness of both intensities, i.e., the trajectory may have only finitely many jumps on any finite interval of time. For establishing convergence rate to the stationary regime, we assume similarly to [18],

\[ \inf_{n>0} h(n, x) \geq \frac{C_0}{1 + x}, \quad x \geq 0. \]

Theorem 2 Let the functions \( \lambda \) and \( h \) be Borel measurable and bounded, the assumption (10) holds and \( C_0 \) satisfies the condition

\[ C_0 > 4(1 + 2\Lambda), \]

and let let the value \( k > 1 \) satisfy

\[ C_0 > 2^{k+1}(1 + \Lambda 2^k). \]
Then there exists a unique stationary measure $\mu$ and there exist $m > k$, $C > 0$ such that for any $t \geq 0$,
\[
\|\mu_t^{n,x} - \mu\|_{TV} \leq C \frac{(1 + n + x)^m}{(1 + t)^{k+1}},
\]
where $\mu_t^{n,x}$ is a marginal distribution of the process $(X_t, t \geq 0)$ with the initial data $X = (n, x) \in X$.

**Remark 1** Under the assumption (13), the inequality (12) holds true, at least, for all values of $k > 1$ close enough to one.

**Remark 2** Existence and uniqueness of invariant measure may be established under weaker assumptions. However, we do not discuss this issue in this paper, as our goal – beside Dynkin’s formula – is only convergence rate, for which we need the condition (11) anyway.

## 3 Auxiliary results

The first Lemma is an $L_1$-version of the Theorem 2.8.2 from [11] from integration theory, originally formulated in $L_2$ as a tool in a Doob’s construction related to stochastic integrals (the latter are not used here). In this paper, the Lemma will serve as a bridge from the standard complete probability formula to its integral analogue.

**Lemma 1** For any $g \in L_1[0,T],$
\[
\int_0^1 \int_0^T |g_{\kappa_m(s)} - g_s| \, ds \, da \to 0,
\]
and for almost all $a \in [0,1],$
\[
\int_0^T |g_{\kappa_m(s)} - g_s| \, ds \to 0,
\]
over some subsequence $m' \to \infty$, where it was denoted
\[
\kappa_m(s) := [2^m(s + a)]2^{-m} - a,
\]
where, in turn, $[a]$ is the integer value of $a \in R$ (i.e., the nearest integer approximation of $a$ from the left). It is accepted that outside $[0,T]$ the function $g$ equals identically zero.
The proof is an exact repetition of the calculus in the proof of \[11, \text{Theorem 2.8.2}\] with a tiny difference of $L_2$ replaced by $L_1$, which does admit this small change. Hence, the details are dropped. For convenience of the reader we only recall that it follows from approximation of $g \in L_1$ by continuous functions, which are dense in $L_1$ and establishing the statement for continuous functions, which is evident. In the proof of the Theorem \[1\] below a similar trick will be used for establishing a similar auxiliary assertion \[30\], in which a one-dimensional nature of time seems to be important.

The second Lemma provides a rigorous proof of the well-known properties that probability of “one event” on a small nonrandom interval of length $\Delta$ is of the order $O(\Delta)$ and probability of “two or more events” on the same interval is of the order $O(\Delta^2)$; although this is a “common knowledge” in queueing theory, the authors believe that for discontinuous intensities it must be justified.

In the next Lemma we accept the following convention: all trajectories of the process are right continuous with left limits; respectively, if the value of the process $X_t$ at $t$ is given, the phrase “no jumps on $(t, t+s]$ relates to no jumps after the moment $t$, but not at $t$, where we are not aware whether or not $X_{t-} = X_t$. At the same time, “no jumps on $(0, s]$” includes no jump at zero, as negative values of $t$ are not allowed, so that automatically $X_0 = X_{0+}$.

**Lemma 2** Under the assumptions of the Theorem\[4\] for any $t \geq 0$,

$$P_{X_t}(\text{no jumps on } (t, t + \Delta)] = \exp \left( - \int_0^\Delta (\lambda + h)(X_t + s) \, ds \right) = 1 + O(\Delta), \quad (14)$$

$$P_{X_t}(\text{at least one jump on } (t, t + \Delta]) = O(\Delta), \quad (15)$$

$$P_{X_t}(\text{exactly one jump up & no down on } (t, t + \Delta]) = \int_0^\Delta \lambda(X_t + s) \, ds + O(\Delta^2), \quad (16)$$

$$P_{X_t}(\text{exactly one jump down & no up on } (t, t + \Delta]) = \int_0^\Delta h(X_t + s) \, ds + O(\Delta^2), \quad (17)$$

and

$$P_{X_t}(\text{at least two jumps on } (t, t + \Delta]) = O(\Delta^2). \quad (18)$$

In all cases above, $O(\Delta)$ and $O(\Delta^2)$ are uniform with respect to $X_t$ and only depend on the norm $\sup_X(\lambda(X) + h(X))$, that is, there exist $C > 0$, $\Delta_0 > 0$ such that for any $X$ and any $\Delta < \Delta_0$,

$$\limsup_{\Delta \to 0} \left\{ \Delta^{-1} P_X(\text{at least one jump on } (0, \Delta]) \right\}$$
\[ + \Delta^{-2} P_X(\text{at least two jumps on } (0, \Delta]) \]

\[ \leq C < \infty. \] (19)

**Proof. A.** The assertions (14), as well as (15), follows from the construction (see (4) and (5), since \( \lambda \) and \( h \) are bounded. (It also follows from the next steps of the proof.)

**B.** Now it is convenient to proceed with (18); it will be used in the proof of (16) and (17). It suffices to consider \( t = 0 \) and any fixed initial value \( X \). Denote

\[ s^N_j = \frac{j \Delta}{2^N}, \quad 1 \leq j \leq 2^N, \quad N \geq 1, \]

and let

\[
B := \{ \text{two or more jumps - up or down - on } (0, \Delta) \},
\]

\[
B^N := \bigcup_j \{ \text{no jumps on } (0, s^N_{j-1}], \text{at least one jump on } (s^N_{j-1}, s^N_j] \}
\]

and at least one jump on \( (s^N_j, \Delta] \},

and

\[ p := P_X(B), \quad p_N := P_X(B^N). \]

Then,

\[ p_N = \sum_{j \leq 2^N} P_X(\text{no jumps on } (0, s^N_{j-1}], \text{at least one jump on } (s^N_{j-1}, s^N_j] \]

and at least one jump on \( (s^N_j, \Delta]) \).

Notice that for any nonrandom \( t \), \( P(X_t \neq X_{t-}) = 0 \). We have,

\[ B^N \subset B, \]

and, moreover,

\[ B \setminus B^N \downarrow \emptyset, \quad N \uparrow \infty. \]

9
The latter holds because for any two jumps there exists \( N \) such that they would be covered by two different intervals, say, \((s_{j-1}^N, s_j^N)\) and \((s_{k-1}^N, s_k^N)\) with \( k \neq j \). In other words, for each \( \omega \in B \) there exists \( N \) such that \( \omega \in B_N^\infty \). Hence, by continuity of a probability measure,

\[
p_N \to p, \quad N \to \infty.
\]

From the Markov property and (14–15), if \( \Delta \) is small enough then we have,

\[
p_N = \sum_{j \leq 2^N} E_X E \left( 1(\text{no jumps on } (0, s_{j-1}^N]) \right.
\]

\[
\times 1(\text{at least one jump on } (s_{j-1}^N, s_j^N]) \mid \mathcal{F}_{s_j^N}^X \times 1(\text{at least one jump on } (s_j^N, \Delta]) \right)
\]

\[
= \sum_{j \leq 2^N} E_X \left( 1(\text{no jumps on } (0, s_{j-1}^N]) \right.
\]

\[
\times 1(\text{at least one jump on } (s_{j-1}^N, s_j^N])) \times E \left( 1(\text{at least one jump on } (s_j^N, \Delta] \mid X_{s_j^N} \right)
\]

\[
= \sum_{j \leq 2^N} E_X \left( 1(\text{no jumps on } (0, s_{j-1}^N]) \right.
\]

\[
\times 1(\text{at least one jump on } (s_{j-1}^N, s_j^N])) \times \left( 1 - \exp \left( - \int_{s_j^N}^\Delta (\lambda + h)(X + s) \, ds \right) \right) \leq C \Delta \sum_{j \leq 2^N} E_X 1(\text{no jumps on } (0, s_{j-1}^N], \text{at least one jump on } (s_{j-1}^N, s_j^N])
\]

\[
= C \Delta P_X (\text{at least one jump up on } (0, s_{N}^N])
\]

\[
= C \Delta \left( 1 - \exp \left( - \int_0^\Delta (\lambda + h)(X + s) \, ds \right) \right) \leq C^2 \Delta^2.
\]

Important is that here the last value of \( C^2 \) is the same for all values of \( N \) and \( X \). Hence, by monotone convergence, we also have

\[
p \leq C^2 \Delta^2,
\]

as required. Notice that (16) and (17) were not used so far.

C. Let us show (16). Informally,

\[
P_{X_t} (\text{exactly one jump up & no down on } (t, t + \Delta])
\]
\[
\begin{align*}
&= \int_0^\Delta \lambda(X_t + s) \exp \left(- \int_0^s \lambda(X_t + r) \, dr \right) \exp \left(- \int_0^\Delta \lambda(X_t^+ + r) \, dr \right) \\
&\quad \times \exp \left(- \int_0^s h(X_t + r) \, dr \right) \exp \left(- \int_0^\Delta h(X_t^+ + r) \, dr \right) \, ds \\
&= \int_0^\Delta \lambda(X_t + s) \, ds + O(\Delta^2),
\end{align*}
\]

as required and with a uniform \(O(\Delta^2)\). Some drawback of this explanation is that it uses an integral identity as if it were complete probability formula, while the latter formula is stated for a split of \(\Omega\) into finitely or countably many disjoint events only. Usually there is no problem with such integration, at least for Riemann integrable functions, since integration is a limit of Darboux sums. However, in our case we assume the integrands, in general, only Lebesgue integrable. So, let us do this estimate more rigorously. We will use the already established formula (18) and instead of (16) we will estimate a slightly different probability of at least one jump up earlier than down on \((0, \Delta]\) (we mean that jump down may or may not occur on this interval). According to our construction, in particular, (6), the value of this probability may be written precisely by using the (conditionally independent given \(X_t\)) densities (7):

\[
\begin{align*}
f^u(z) &:= \lambda(X_t + z) \exp \left(- \int_0^z \lambda(X_t + s) \, ds \right) \\

f^d(z) &:= h(X_t + z) \exp \left(- \int_0^z h(X_t + s) \, ds \right).
\end{align*}
\]

We have,

\[
P_X(\text{at least one jump up earlier than down on } (0, \Delta])
\]

\[
= \int_0^\infty \int_0^\infty 1(z^1 < z^2)1(z^1 \leq \Delta)f^u(z^1)f^d(z^2) \, dz^1 \, dz^2
\]

\[
\leq \int_0^\infty \int_0^\infty 1(0 < z^2)1(z^1 \leq \Delta)f^u(z^1)f^d(z^2) \, dz^1 \, dz^2
\]

\[
= \int_0^\Delta f^u(z^1) \, dz^1 = \int_0^\Delta \lambda(X + z) \exp \left(- \int_0^z \lambda(X + s) \, ds \right) \, dz
\]
\[
\leq \int_0^\Delta \lambda (X + z) \, dz.
\]

On the other hand,

\[
P_X(\text{at least one jump up earlier than down on } (0, \Delta])
\]

\[
= \int_0^\infty \int_0^\infty 1(z_1 < z^2)1(z_1 \leq \Delta) f^u(z_1) f^d(z_2) \, dz_1 \, dz_2
\]

\[
\geq \int_0^\Delta \int_0^\infty 1(z_1 \leq \Delta) f^u(z_1) f^d(z_2) \, dz_1 \, dz_2
\]

\[
= \int_0^\Delta f^u(z_1) \, dz_1 \times \int_0^\Delta f^d(z_2) \, dz_2
\]

\[
\geq \int_0^\Delta \lambda (X + z) \exp \left( -\int_0^\Delta \lambda (X + s) \, ds \right) \, dz \times \exp \left( -\int_0^\Delta h(X + s) \, ds \right)
\]

\[
= \int_0^\Delta \lambda (X + z) \, dz \times \exp \left( -\int_0^\Delta (\lambda + h)(X + s) \, ds \right)
\]

\[
= (1 + O(\Delta)) \int_0^\Delta \lambda (X + z) \, dz = \int_0^\Delta \lambda (X + z) \, dz + O(\Delta^2),
\]

with uniform \(O(\Delta)\) and \(O(\Delta^2)\), as required. Since we already know that probability of two or more jumps is of the order \(\Delta^2\) – see (18) – this justifies (16).

D. The statement (17) is established similarly. Note that in the previous steps, expressions like \(1 - \exp \left( -\int_0^\Delta \psi(t) \, dt \right)\) with some non-negative bounded \(\psi(t)\) are evaluated, say, \(\psi(t) < M, \forall t\). In this case, this exponent may be expanded via Taylor’s series of the variable \(\int_0^\Delta \psi(t) \, dt\). For \(\Delta < 2/M\) this series satisfies the well-known condition for the alternating series test, which implies strict bounds
1 - \exp\left(-\int_0^\Delta \psi(t) \, dt\right) < M\Delta \text{ and } 0 < \left(1 - \exp\left(-\int_0^\Delta \psi(t) \, dt\right)\right) - \int_0^\Delta \psi(t) \, dt < M^2\Delta^2/2. \text{ This shows that the estimates in the proof above and in (19) are uniform, as required. The Lemma 2 is proved.}

**Lemma 3** Under the assumptions of the Theorem 1, for any \( f \in C_b(\mathcal{X}) \) the function \( T_t f(X) = E_X f(X_t) \) is continuous in \( t \).

**Proof.** It suffices for any \( f \in C_b(\mathcal{X}) \) and for small \( \Delta > 0 \) to show that for any \( t \geq 0 \),

\[ |E_X(f(X_{t+\Delta}) - f(X_t))| \to 0, \quad \Delta \to 0. \quad \text{(20)} \]

and for any \( t > 0 \),

\[ |E_X(f(X_{t-\Delta}) - f(X_t))| \to 0, \quad \Delta \to 0. \quad \text{(21)} \]

The convergence (20) follows straightforward from the first statement of the Lemma 2 by virtue of Lebesgue’s dominated convergence theorem, due to the cadlag property of the process. Similarly, we have,

\[ |E_X(f(X_{t-\Delta}) - f(X_{t-}))| \to 0, \quad \Delta \to 0. \]

However, by construction (see (4)), jump at any nonrandom time \( t \) has probability zero, that is,

\[ P(X_t = X_{t-}) = 1. \]

This implies (21). The Lemma 3 is proved.

Denote \( X' = (n',x') \uparrow X = (n,x) \) iff \( x' \uparrow x \) and \( n' = n \) when \( x' \) is close enough to \( x \). Similarly, \( X' = (n',x') \downarrow X = (n,x) \) iff \( x' \downarrow x \) and \( n' = n \) for \( x' \) close enough to \( x \).

**Lemma 4** Under the assumptions of the Theorem 1 the process \( (X_t, t \geq 0) \) is Feller, that is, \( T_t f(\cdot) \in C_b(\mathcal{X}) \) for any \( f \in C_b(\mathcal{X}) \).

**Proof.** A. Denote \( \phi(t,X) := E_X f(X_t) \). Let \( (n',x') = X' \uparrow X = (n,x) \). Without loss of generality we may assume \( n' \equiv n \) and \( x' \uparrow x \). Let \( s := x - x' > 0 \). Note that for the process with initial data \( X' = (n',x') \),

\[ x_s1(\text{no jumps on } (0,s]) = x1(\text{no jumps on } (0,s]). \]

So, for \( x' \) close enough to \( x \), by the Lemma 2,

\[ \phi(t,X') = E_{X'} f(X_t) \]

\[ = E_{X'} f(X_t) \mathbb{1}(\text{no jumps on } (0,s]) \]
\[ +E_X f(X_t) 1(\text{at least one jump on } (0, s]) \]
\[ = E_X 1(\text{no jumps on } (0, s]) E_X f(X_{t-s}) + O(s) \]
\[ = E_X 1(\text{no jumps on } (0, s]) E_X f(X_{t-s}) + O(s) \]
\[ = E_X 1(\text{no jumps on } (0, s]) \phi(t - s, X) + O(s) \]
\[ = (1 - O(s)) \phi(t - s, X) + O(s). \]

By virtue of the Lemma 3 and due to our notation \( s = x - x' \),
\[ \lim_{x' \uparrow x} \phi(t, X') = \lim_{s \downarrow 0} \phi(t - s, X) = \phi(t, X). \]

**B.** Now let \( X' \downarrow X \). In this case denote \( s := x' - x \ (s > 0) \). Then similarly to the above,
\[ E_X f(X_t) = O(s) + (1 - O(s)) E_X f(X_{t-s}) \]
Hence, we have,
\[ \phi(t, X) = \phi(t - s, X') + O(s). \]
Since by the Lemma 3,
\[ \phi(t, X') = \phi(t - s, X') + o(1), \]
we obtain,
\[ \phi(t, X') = \phi(t - s, X') + o(1) = \phi(t, X) + O(s) + o(1), \]
or, equivalently,
\[ \lim_{X' \downarrow X} \phi(t, X') = \phi(t, X). \]
This completes the proof of the Lemma 4.

**Remark 3** Note that in [5] Feller’s property is proved for a more general model, however, under the additional condition that intensities are continuous. In our case continuity is not necessary – which is in line with the very idea of construction of a process with discontinuous intensities – and in the sequel this Feller’s property will be used for establishing strong Markov property by using classical tools. On the other hand, in [5] strong Markov property is established independently from Feller’s one.
4 Proof of Theorem

As was already mentioned above, both formulae (1) and (2) are, actually, versions of complete probability formula (it would be better to say, complete expectation). If it were possible to replace the integrals by series, they both would have been, indeed, complete probability (expectation) formulae as suggested. However, in our case, integrals and series are not the same things. Notice that for continuous and bounded \( \lambda \) and \( h \), the formula (1) with any \( f \in C^1_b(X) \) is, indeed, a simple corollary of a standard analysis of probabilities of transitions over a small time, as in the derivation of the Kolmogorov forward equations. This is why \( \mathcal{G} \) may be regarded as a natural candidate to a generalised generator for discontinuous intensities, too.

**Ia.** In the first part of the proof we assume \( f \in C^1_0(X) \), rather than \( f \in C^1_b(X) \) in the formula (1). In some lines it is useful to know that \( f \) and its derivative are uniformly continuous. In the end of the proof, the assertion (1) will be extended from \( C^1_0(X) \) to the whole \( C^1_b(X) \). Let \( \Delta := 2^{-m}t \),

\[
t_0 = 0, \quad t_i = \kappa_m \left( \frac{it}{2^m} \right) \quad \forall \, 0 \equiv ([it + a2^m]2^{-m} - a) \quad \forall \, 0 \leq i \leq 2^m, \quad t_{2^m+1} = t.
\]

Let \( \Delta_i := t_{i+1} - t_i \). Notice that \( \Delta_i = \Delta, \, 1 \leq i \leq 2^m \) and \( \Delta_i \leq \Delta \) for any \( i \).

Consider the difference

\[
E_X f(X_t) - f(X) = E_X \sum_{i=0}^{2^m} E_{X_{t_i}}(f(X_{t_{i+1}}) - f(X_{t_i})).
\]

Consider one term from this sum. Emphasize that they are all treated similarly. This term may be split into four parts:

\[
E_{X_{t_i}}(f(X_{t_{i+1}}) - f(X_{t_i})) = E_{X_{t_i}}(f(X_{t_{i+1}}) - f(X_{t_i}))1(\text{no jumps on } (t_i, t_{i+1}])
+ E_{X_{t_i}}(f(X_{t_{i+1}}) - f(X_{t_i}))1(\text{one jump up & no down on } (t_i, t_{i+1}])
+ E_{X_{t_i}}(f(X_{t_{i+1}}) - f(X_{t_i}))1(\text{one jump down & no up on } (t_i, t_{i+1}])
+ E_{X_{t_i}}(f(X_{t_{i+1}}) - f(X_{t_i}))1(\text{at least, two jumps on } (t_i, t_{i+1}])
\equiv \mathcal{I}_1 + \ldots + \mathcal{I}_4.
\]

**Ib.** We have, with notation \( X + s := (n, x + s) \) for any \( X = (n, x) \),

\[
\mathcal{I}_i = \exp \left( - \int_0^{\Delta_i} (\lambda(X_{t_i} + s) + h(X_{t_i} + s)) \, ds \right) (f(X_{t_i} + \Delta_i) - f(X_{t_i})),
\]

15
\[
= (1 + O(\Delta)) \int_0^{\Delta} \frac{\partial}{\partial x} f(X_{t_i} + s) \, ds.
\]

Here it is likely that we could use the Lemma \[\text{1}\] but we prefer easier methods where possible. Notice that

\[
\int_0^{\Delta} \frac{\partial}{\partial x} f(X_{t_i} + s) \, ds = \int_0^{\Delta} \frac{\partial}{\partial x} f(X_{t_i+s}) \, ds
\]

(22)

(23)

After summation and taking expectation, this term gives us

\[
E_X \sum_{i=0}^{2^m} I_i = E_X \sum_{i=0}^{2^m} \int_0^{\Delta} \frac{\partial}{\partial x} f(X_{t_i} + s) \, ds = E_X \sum_{i=0}^{2^m} \int_0^{\Delta} \frac{\partial}{\partial x} f(X_{t_i+s}) \, ds
\]

\[
+ E_X \sum_{i=0}^{2^m} \int_0^{\Delta} \left( \frac{\partial}{\partial x} f(X_{t_i} + s) - \frac{\partial}{\partial x} f(X_{t_i+s}) \right) \, ds.
\]

By virtue of the cadlag property of \((X_s, s \geq 0)\) and \(f \in C^1_b\), we have by Lebesgue’s dominated convergence Theorem,

\[
E_X \sum_{i=0}^{2^m} \int_0^{\Delta} \left( \frac{\partial}{\partial x} f(X_{t_i} + s) - \frac{\partial}{\partial x} f(X_{t_i+s}) \right) \, ds
\]

\[
= E_X \int_0^t \left( \frac{\partial}{\partial x} f(X_{\kappa_m^s(s)} + s - \kappa_m^a(s) - \frac{\partial}{\partial x} f(X_s) \right) \, ds
\]

\[
= E_X \int_0^t \left( \frac{\partial}{\partial x} f(X_{\kappa_m^s(s)} + s - \kappa_m^a(s) - \frac{\partial}{\partial x} f(X_s) \right) \, ds \to 0, \quad \Delta \to 0,
\]

the latter equality a.s., because

\[P_X(X_{\kappa_m^s(s)} + s - \kappa_m^a(s) - X_{s_-} \to 0, \quad m \to \infty) = 1.\]
Hence – and due to $\int_0^t |f(X_s) - f(X_{s-})| \, ds = 0$ a.s. – we get,

$$E_X \sum_{i=0}^{\infty} I_i^i = E_X \int_0^t \frac{\partial}{\partial x} f(X_s) \, ds + o(1), \quad \Delta \to 0.$$ 

Ic. Further,

$$I_2^i = E_{X_{t_i}} (f(X_{t_{i+1}}) - f(X_{t_i})) 1(\text{one jump up & no down on } (t_i, t_{i+1}])$$

$$\equiv E_{X_{t_i}} (f(X_{t_i}^+) - f(X_{t_i})) 1(\text{one jump up & no down on } (t_i, t_{i+1}])$$

$$+ E_{X_{t_i}} (f(X_{t_{i+1}}) - f(X_{t_i}^+)) 1(\text{one jump up & no down on } (t_i, t_{i+1}])$$

$$= (f(X_{t_i}^+) - f(X_{t_i})) \left( \int_0^{\Delta_i} \lambda(X_{t_i} + s) \, ds + O(\Delta^2) \right)$$

$$+ E_{X_{t_i}} (f(X_{t_{i+1}}) - f(X_{t_i}^+)) 1(\text{one jump up & no down on } (t_i, t_{i+1}]),$$

by the Lemma 2. After summation and taking expectation, the first term here gives

$$E_X \int_0^t (f(X_{t_i}^+ - f(X_{t_i})) \lambda(X_{t_i}^+ + s - \kappa^a(s)) \, ds + O(\Delta),$$

and we will explain below – via the Lemma 1 – why it approaches the desired

$$E_X \int_0^t (f(X_{s-}) - f(X_{s-})) \lambda(X_s) \, ds = E_X \int_0^t (f(X_s^+) - f(X_s)) \lambda(X_s) \, ds, \quad \text{a.s.}$$

The absolute value of the second term in (24) does not exceed

$$\rho_f(\Delta) E_{X_{t_i}} 1(\text{one jump up & no down on } (t_i, t_{i+1}]),$$

where $\rho_f$ is the modulus of continuity of the function $f$. By virtue of the Lemma 2 (see (16)), after summation and expectation this gives us

$$\rho_f(\Delta) E_X \int_0^t \lambda(X_{\kappa^a_m(s)} + s - \kappa^a_m(s)) \, ds + O(\Delta) = o(1), \quad \Delta \to 0.$$
**Id.** Similarly with some \( \theta \in (0, \Delta_i) \),

\[
I_3^i = (1 - \exp \left( - \int_0^{\Delta_i} h(X_{t_i} + s) \, ds \right) + O(\Delta^2)) (f(X^-_{t_i} + \theta) - f(X_{t_i}))
\]

\[
= \left( \int_0^{\Delta_i} h(X_{t_i} + s) \, ds + O(\Delta^2) \right) (f(X^-_{t_i}) - f(X_{t_i}) + \rho_f(\Delta)).
\]

Since \( \rho_f(\Delta) = o(1) \), we get,

\[
E_x \sum_{i=0}^{2m} I_3^i =
\]

\[
= E_x \int_0^t h(X_{\kappa_m(x)} + s - \kappa_m(s)) \, ds (f(X^-_{\kappa_m(x)}) - f(X_{\kappa_m(x)}) + o(1)).
\]

It will be shown below that the main term here approaches

\[
E_x \int_0^t h(X_s)(f(X^-_s) - f(X_s)) \, ds \overset{a.s.}{=} E_x \int_0^t h(X_s)(f(X^-_s) - f(X_s)) \, ds.
\]

**Ie.** Finally (see the Lemma 2), with uniform \( O(\Delta^2) \),

\[
I_4^i = O(\Delta^2) \Rightarrow E_x \sum_{i=0}^{2m} I_4^i = O(\Delta).
\]

**If.** Now in the expression arising from the terms \( E_x I_2^i \), as well as from \( E_x I_3^i \), we would like to replace \( f(X^+_{\kappa_m(x)}) - f(X_{\kappa_m(x)}) \) by \( f(X^+_s) - f(X_s) \) and \( f(X^-_{\kappa_m(x)}) - f(X_{\kappa_m(x)}) \) by \( f(X^-_s) - f(X_s) \), respectively; also, \( h(X_{\kappa_m(x)}) \) and \( \lambda(X_{\kappa_m(x)}) \) will be replaced by \( h(X_s) \) and \( \lambda(X_s) \), respectively. By virtue of continuity of \( f \) and cadlag property of the process, we estimate for almost every \( a \in [0, 1] \),

\[
E_x \int_0^t \lambda(X_{\kappa_m(x)}) (f(X^+_{\kappa_m(x)}) - f(X_{\kappa_m(x)})) \, ds
\]

\[
- E_x \int_0^t \lambda(X_s)(f(X^+_s) - f(X_s)) \, ds \rightarrow 0, \quad m \rightarrow \infty,
\]
and

\[
\left( E_X \int_0^t h(X_{\kappa_{m'}^a(s)}) (f(X_{\kappa_{m'}^a(s)}^-) - f(X_{\kappa_{m'}^a(s)}) \right) ds - \\
- E_X \int_0^t h(X_s) (f(X_{s+}) - f(X_s)) ds \right) \to 0, \quad m' \to \infty,
\]

where \( m' \to \infty \) is some subsequence guaranteed by the Lemma 1. Indeed, say, for \( h \), by the Lemma 1 for a.e. \( a \),

\[
E_X \int_0^t \left| (h(X_{\kappa_{m'}^a(s)}) - h(X_s)) (f(X_{\kappa_{m'}^a(s)}^+) - f(X_{\kappa_{m'}^a(s)}^-)) \right| ds \\
\leq C_f E_X \int_0^t \left| (h(X_{\kappa_{m'}^a(s)}) - h(X_s)) \right| ds \to 0, \quad m' \to \infty.
\]

Further, by continuity of \( f \) and cadlag property of \( X \),

\[
\left( E_X \int_0^t h(X_s) (f(X_{\kappa_{m'}^a(s)}^+) - f(X_{\kappa_{m'}^a(s)}^-)) ds - \\
- E_X \int_0^t h(X_s) (f(X_{s+}) - f(X_{s-})) ds \right) \to 0, \quad m' \to \infty,
\]

by virtue of the Lebesgue dominated (bounded) convergence theorem, as \( f(X_s) \) is cadlag, both \( h \) and \( f \) are bounded and

\[
\kappa_{m'}^a(s) \to s-, \quad m' \to \infty.
\]

Finally, since \( P(X_s \neq X_{s-}) = 0, \forall s, \) we get,

\[
E_X \left| \int_0^t h(X_s) (f(X_{s+}^-) - f(X_{s-})) ds - \int_0^t h(X_s) (f(X_{s+}) - f(X_s)) ds \right| = 0.
\]

**Ig.** Further, we would like to have \( \lambda(X_s) \) under the integral instead of \( \lambda(X_{t_i} + (s - t_i)) \) in each term (24)

\[
(f(X_{t_i}^+)) \left( \int_0^{\Delta_i} \lambda(X_{t_i} + s) ds + O(\Delta^2) \right).
\]

19
So, let us estimate the difference

\[
E_X \sum_{i=0}^{2^m} \left( \int_{t_{i+1}}^{t_{i+1}} \lambda(X_{t_i} + (s - t_i)) \, ds - \int_{t_i}^{t_{i+1}} \lambda(X_s) \, ds \right) (f(X_{t_i}^+) - f(X_{t_i}))
\]

Recall that \( t_i = \kappa_m^a \left( \frac{it}{2^m} \right) \). Let us show that for almost any \( a \in [0, 1] \),

\[
E_X \sum_{i=0}^{2^m} \left| \int_{t_i}^{t_{i+1}} \lambda(X_{t_i} + (s - t_i)) \, ds - \int_{t_i}^{t_{i+1}} \lambda(X_s) \, ds \right| |f(X_{t_i}^+) - f(X_{t_i})| = 0,
\]

(25)

as \( \Delta \to 0 \).

Firstly, according to the Lemma 1 there exists a subsequence \( (m' \to \infty) \) such that for almost every \( a \in [0, 1] \)

\[
\int_0^t |\lambda(X_s) - \lambda(X_{\kappa_m^a(s)})| \, ds \to 0,
\]

and for the whole sequence \( (m \to \infty) \),

\[
\int_0^1 \int_0^t |\lambda(X_s) - \lambda(X_{\kappa_m^a(s)})| \, ds \, da \to 0.
\]

(26)

Now let us show that

\[
\int_0^1 \int_0^t |\lambda(X_{\kappa_m^a(s)} + (s - \kappa_m^a(s))) - \lambda(X_{\kappa_m^a(s)})| \, ds \, da \to 0, \ m \to \infty.
\]

(27)

The idea is very similar to that of the Lemma 1 however, here a direct reference to this Lemma is questionable and, hence, we need to provide an independent proof. Recall that the trajectory \((X_s)\) may have only finitely many jumps on \((0, t]\). Also, except for the points of jumps, the evolution of \((X)\) is deterministic and linear with a constant positive speed. In other words, integration of the composite function \(\lambda(X)\) turns out to be equivalent to the integration of \(\lambda(\cdot)\) over some interval between two consequent jumps of the process \((X)\). This allows us to use the same trick from the Lemma 1 but this time we approximate the function \(\lambda(\cdot)\) rather than the composite
function $\lambda(X)$. Emphasize that this is possible exactly because of the piecewise linear law of evolution of $(X_s)$ between the consequent moments of jumps and because there are only finitely many jumps on each trajectory.

So, let us approximate the function $\lambda(\cdot)$ by bounded uniformly continuous functions — say, $\lambda^\epsilon$ — in the topology of convergence in $L_1[0,T]$ for every $T > 0$. Note that it is not enough to approximate $\lambda(\cdot)$ on $(0,t]$. Denote by $(\tau_j, j = 0, 1, \ldots)$ the sequence of moments of jumps of the trajectory of $X$; recall that for each $\omega$ there are only finitely many of them on $(0,t]$.

We have, on each $(\tau_j, \tau_{j+1}]$ and for every $m$,

$$\int_0^{\tau_{j+1}} \int_0^{\tau_j} |\lambda^\epsilon(X_{\kappa_m^a(s)}) - \lambda(X_{\kappa_m^a(s)})| \, ds \, da = \int_0^{\tau_{j+1}} \int_0^{\tau_j} |\lambda^\epsilon(X_{\kappa_m^a(s)}) - \lambda(X_{\kappa_m^a(s)})| \, da \, ds$$

$$\leq \int_{\tau_j}^{\tau_{j+1}} 2^m \|\lambda^\epsilon - \lambda\|_{L_1([\tau_j - 2^{-m}, \tau_{j+1}])} \, ds = \|\lambda^\epsilon - \lambda\|_{L_1([\tau_j - 2^{-m}, \tau_{j+1}])} \to 0, \epsilon \to 0.$$

Since the number of moments $\tau_j$ on $(0,t]$ is finite for (almost) all $\omega$, it also follows for each $m$ that

$$\int_0^t \int_0^t |\lambda^\epsilon(X_{\kappa_m^a(s)}) - \lambda(X_{\kappa_m^a(s)})| \, ds \, da$$

$$= \int_0^{\tau_{j+1}} \sum_j \int_{\tau_j}^{\tau_{j+1}} |\lambda^\epsilon(X_{\kappa_m^a(s)}) - \lambda(X_{\kappa_m^a(s)})| \, ds \, da \to 0, \epsilon \to 0. \tag{28}$$

Now consider the difference

$$\int_0^t \int_0^t |\lambda^\epsilon(X_{\kappa_m^a(s)} + (s - \kappa_m^a(s))) - \lambda(X_{\kappa_m^a(s)} + (s - \kappa_m^a(s)))| \, ds \, da.$$

On each $(\tau_j, \tau_{j+1}]$ and for every $m$, due to integration over $s$,

$$\int_0^{\tau_{j+1}} \int_0^{\tau_j} |\lambda^\epsilon(X_{\kappa_m^a(s)} + (s - \kappa_m^a(s))) - \lambda(X_{\kappa_m^a(s)} + (s - \kappa_m^a(s)))| \, ds \, da$$

$$\leq \int_0^{\tau_{j+1}} \|\lambda^\epsilon - \lambda\|_{L_1([\tau_j - 2^{-m}, \tau_{j+1}])} \, da = \|\lambda^\epsilon - \lambda\|_{L_1([\tau_j - 2^{-m}, \tau_{j+1}])} \to 0, \epsilon \to 0.$$
Therefore, also

\[
\left(\int_0^1 \int_0^t \left| \lambda^\varepsilon(X_{\kappa_m^\varepsilon(s)} + (s - \kappa_m^\varepsilon(s))) - \lambda(X_{\kappa_m^\varepsilon(s)} + (s - \kappa_m^\varepsilon(s))) \right| ds \, da \right) \to 0, \quad \varepsilon \to 0.
\] (29)

Notice that by virtue of continuity of \(\lambda^\varepsilon\),

\[
\int_0^1 \int_0^t \left| \lambda^\varepsilon(X_{\kappa_m^\varepsilon(s)} + (s - \kappa_m^\varepsilon(s))) - \lambda^\varepsilon(X_{\kappa_m^\varepsilon(s)}) \right| ds \, da \to 0, \quad m \to \infty,
\] (30)

for each \(\omega\) and \(\varepsilon\). So, due to (28), (29) and (30), the left hand side in (27) goes to zero as \(m \to \infty\), as it does not depend on \(\varepsilon\). From here it easily follows that there exists a subsequence \((m')\) such that for almost every \(a\),

\[
\int_0^t \left| \lambda(X_{\kappa_{m'}^{a}(s)} + (s - \kappa_{m'}^{a}(s))) - \lambda(X_{\kappa_{m'}^{a}(s)}) \right| ds \to 0, \quad m' \to \infty.
\] (31)

Moreover, this subsequence may be chosen from the earlier fixed subsequence for which (26) is valid. Hence, we may assume that for \((m')\) both (26) and (31) are valid simultaneously.

Overall, by virtue of Lebesgue’s dominated convergence, we conclude that by virtue of (26) and (31), convergence (25) holds true with almost any \(a \in [0, 1]\), that is, over some subsequence \((m' \to \infty)\),

\[
EX \sum_{i=0}^{m'} \int_{t_i}^{t_{i+1}} \left| \lambda(X_{\kappa_m^{a'}(s)}) - \lambda(X_s) \right| ds \left| f(X_{t_i}^+) - f(X_{t_i}) \right| \to 0.
\] (32)

Similarly, without loss of generality, we may assume that over the same subsequence,

\[
EX \sum_{i=0}^{m'} \int_{t_i}^{t_{i+1}} \left| h(X_{\kappa_m^{a'}(s)}) - h(X_s) \right| ds \left| f(X_{t_i}^-) - f(X_{t_i}) \right| \to 0.
\] (33)

From (32), (33), the statement (1) (Dynkin’s formula) for \(f \in C_0^1(\mathcal{X})\) follows.
II. Now assume \( f \in C^1_b(\mathcal{X}) \). Let us approximate this function by a uniformly bounded sequence \( f^N \in C^1_b(\mathcal{X}) \), so that
\[
\|f^N - f\|_{C^1(K)} \to 0,
\]
for any compact set \( K \in \mathcal{X} \).

Then, by the part I of the proof, we have
\[
E_X f^N(X_t) - f^N(X) = E_X \int_0^t \mathcal{G} f^N(X_s) \, ds,
\]
and we need to justify the passage to the limit as \( N \to \infty \). This follows from Lebesgue’s dominated convergence Theorem. Indeed, for each \( s, t \) and \( \omega \),
\[
f^N(X_t) \to f(X_t), \quad \mathcal{G} f^N(X_s) \to \mathcal{G} f(X_s),
\]
and \( f^N(X_t) \) and \( \mathcal{G} f^N(X_s) \) are uniformly bounded. Hence, the statement of the Theorem \([1]\) about Dynkin’s formula \([1]\) for any \( f \in C^1_b(\mathcal{X}) \) is proved.

III. Now let us show the formula \([2]\). Here we have to consider the difference
\[
E_X \varphi(t, X_t) - \varphi(0, X) = E_X \sum_{i=0}^{2m} E_{X_{t_i}} (\varphi(t_{i+1}, X_{t_{i+1}}) - \varphi(t_i, X_{t_i})).
\]
Consider one term from this sum. This term may be split into five parts,
\[
E_{X_{t_i}} (\varphi(t_{i+1}, X_{t_{i+1}}) - \varphi(t_i, X_{t_i}))
= E_{X_{t_i}} (\varphi(t_i, X_{t_{i+1}}) - \varphi(t_i, X_{t_i})) \mathbb{1} (\text{no jumps on } (t_i, t_{i+1}])
\]
\[
\quad + E_{X_{t_i}} (\varphi(t_i, X_{t_{i+1}}) - \varphi(t_i, X_{t_i})) \mathbb{1} (\text{one jump up on } (t_i, t_{i+1}])
\]
\[
\quad + E_{X_{t_i}} (\varphi(t_i, X_{t_{i+1}}) - \varphi(t_i, X_{t_i})) \mathbb{1} (\text{one jump down on } (t_i, t_{i+1}])
\]
\[
\quad + E_{X_{t_i}} (\varphi(t_i, X_{t_{i+1}}) - \varphi(t_i, X_{t_i})) \mathbb{1} (\text{at least, two jumps on } (t_i, t_{i+1}])
\]
\[
\quad + E_{X_{t_i}} (\varphi(t_{i+1}, X_{t_{i+1}}) - \varphi(t_i, X_{t_{i+1}})).
\]
All terms but the last one are considered similarly to the case \((f(X))\). The last term, clearly, gives us
\[
E_{X_{t_i}} (\varphi(t_{i+1}, X_{t_{i+1}}) - \varphi(t_i, X_{t_{i+1}})) = E_{X_{t_i}} \int_{t_i}^{t_{i+1}} \frac{\partial}{\partial s} \varphi(s, X_{t_{i+1}}) \, ds,
\]
23
which after summation and taking expectation converges as follows,

\[ E_X \sum_{i=0}^{2m} E_{X_{t_i}} \int_{t_i}^{t_{i+1}} \frac{\partial}{\partial s} \varphi(s, X_{t_{i+1}}) \, ds \to E_X \int_0^t \frac{\partial}{\partial s} \varphi(s, X_s) \, ds, \quad m \to \infty, \]

by Lebesgue’s dominated convergence Theorem for any \( \varphi \in C^1_0([0, \infty) \times \mathcal{X}) \), as required. Now we may repeat the arguments about extending the formula from \( C^1_0([0, \infty) \times \mathcal{X}) \) to \( C^1_b([0, \infty) \times \mathcal{X}) \).

IV. It remains to show strong Markov property. This follows from continuity of the function \( E_X f(X_t) \) in \( X \) by virtue of the Lemma 3 and due to the cadlag property of the process, see [6, Theorem 3.4]. The Theorem 1 is proved.

5 Proof of Corollary 1

Notice that the issue is to extend both Dynkin’s formulae to polynomially growing functions. First of all, recall that the formulae (1) and (2) for \( f \in C^1_b(\mathcal{X}) \) and \( \varphi \in C^1_b([0, \infty) \times \mathcal{X}) \) in martingale language read as follows: the processes

\[ M_t := f(X_t) - f(X) - \int_0^t Gf(X_s) \, ds, \quad t \geq 0, \quad (34) \]

and

\[ \tilde{M}_t := \varphi(t, X_t) - \varphi(0, X) - \int_0^t \left( \frac{\partial}{\partial s} \varphi(s, X_s) + G\varphi(s, X_s) \right) \, ds, \quad t \geq 0, \quad (35) \]

are both martingales (cf. [13]). For the reader’s convenience let us show this, say, for \( \tilde{M} \). Indeed,

\[ E|\tilde{M}_t| < \infty \]

since all terms in the right hand side in (35) are bounded for \( \varphi \in C^1_b \), and

\[ E_X(\tilde{M}_t - \tilde{M}_s | \mathcal{F}^X_s) = E_X(\tilde{M}_{t-s} - \tilde{M}_0) = E_X(\tilde{M}_{t-s})|_{X=X_s}, \]

due to Markov property. But due to (2), for \( \text{every} \) initial data \( X \),

\[ E_X(\tilde{M}_{t-s}) = 0 \]

for every \( X \). Hence,

\[ E_X(\tilde{M}_{t-s})|_{X=X_s} = 0. \]

Therefore, \( \tilde{M} \) is, indeed, a (cadlag) martingale.
Now the task is to extend (35) to the function \(L_{k,m}\), or, equivalently, to extend (2) to such function. It is easy to see that \(L_{k,m}\) may be approximated by functions \(\varphi^N \in C^{1}_{b}([0, \infty) \times \mathcal{X})\) so that

\[
\|\varphi^N(\cdot, \cdot) - L_{k,m}(\cdot, \cdot)\|_{C^{1}([0,N] \times [0,N])} \to 0, \quad N \to \infty,
\]

for any \(t \geq 0, X \in \mathcal{X}\), and

\[
\sup_{N} \left( |\varphi^N(t, X)| + \left| \frac{\partial}{\partial x} \varphi^N(t, X) \right| + \left| \frac{\partial}{\partial t} \varphi^N(t, X) \right| \right) \leq C(1 + L_{k,m}(t, X)). \tag{36}
\]

Then it is possible to pass to the limit in all terms of the equation (2) written for \(\varphi^N\), if the following a priori bound is established,

\[
\sup_{t \leq T} E_{(n_0, x_0)}(n_t + x_t)^m \leq C(T, n_0, x_0, m) < \infty, \tag{37}
\]

for any \(m > 0\) with some function \(C(T, n, x, m)\), for any \(T > 0\). Indeed,

\[
\varphi^N(t, X_t) \to \varphi(t, X_t), \quad \frac{\partial}{\partial s} \varphi^N(s, X_s) \to \frac{\partial}{\partial s} \varphi(s, X_s), \quad N \to \infty,
\]

and

\[
\mathcal{G} \varphi^N(s, X_s) \to \mathcal{G} \varphi(s, X_s), \quad N \to \infty.
\]

Since (37) will be established for any \(m > 0\), we obtain by Lebesgue’s convergence Theorem under the uniform integrability condition,

\[
\varphi^N(t, X_t) - \varphi^N(0, X) - \int_{0}^{t} \left( \frac{\partial}{\partial s} \varphi^N(s, X_s) + \mathcal{G} \varphi^N(s, X_s) \right) \, ds
\]

\[
\to \varphi(t, X_t) - \varphi(0, X) - \int_{0}^{t} \left( \frac{\partial}{\partial s} \varphi(s, X_s) + \mathcal{G} \varphi(s, X_s) \right) \, ds, \quad N \to \infty.
\]

This would also imply that \(E_{X}|\tilde{M}_t| + E_{X}|M_t| < \infty\), for \(f = L_m\) and \(\varphi = L_{k,m}\).

The easiest explanation of (37) is, apparently, to use the fact that the process \((X_t = (n_t, x_t), t \geq 0)\) is dominated by a similar process, say, \((\bar{X}_t = (\bar{n}_t, \bar{x}_t), t \geq 0)\) without actually serving the customer at the server and with a constant arrival rate \(\bar{\lambda} := \Lambda \vee \lambda_0\), that is, for each \(\omega\),

\[
\bar{x}_t \geq x_t, \quad \bar{n}_t \geq n_t, \quad t \geq 0.
\]

For the process \((\bar{X})\) we have,

\[
\bar{x}_t = x + t
\]

and

\[
\bar{n}_t = n_0 + \xi_t
\]
where $\xi_t$ has a Poisson distribution with parameter $\bar{\lambda}t$. We may imagine a situation as if the current serving has suddenly pended and the customer remained at the idle server forever. We have,

$$
\psi(t, m) := E\xi_t^m = \sum_{j=0}^{\infty} j^m \frac{(\bar{\lambda}t)^j}{j!} e^{-\bar{\lambda}t} < \infty,
$$

so,

$$
\sup_{t \leq T} E(n_0, x_0)(\bar{n}_t + \bar{x}_t)^m \leq 3^{m-1}(x_0 + T)^m + 3^{m-1}n_0^m + 3^{m-1}\psi(T, m).
$$

Hence, (37) holds true and the statement of the Corollary [1] is proved.

6 Proof of Theorem [2]

The proof repeats the calculus in [18] based on Lyapunov functions

$$
L_m(X) := (n + 1 + x)^m \quad \text{and} \quad L_{m,k}(t, X) := (1 + t)^k(n + 1 + x)^m
$$

and on Dynkin’s formulae (11) and (2) due to the Corollary [1]. The news is only a wider class of intensities, which may be discontinuous, however, this does not affect the calculus at all once (11) and (2) are established. Hence, we drop the details.

References

[1] Asmussen, S., Applied Probability and Queues, 2nd edition. Springer, Berlin et al., 2003.

[2] Bambos, N., Walrand, J., On stability of state-dependent queues and ayclic queueing networks // Adv. Appl. Probab. 1989, 21(3), 681–701.

[3] Borovkov, A. A., Boxma, O. J., Palmowski, Z., On the Integral of the Workload Process of the Single Server Queue // Journal of Applied Probability, 2003, 40(1), 200–225.

[4] Bramson, M., Stability of Queueing Networks: École d’Été de Probabilités de Saint-Flour XXXVI-2006, Lecture Notes in Math., 2008, Vol. 1950.

[5] Davis, M. H. A., Piecewise-Deterministic Markov Processes: A General Class of Non-Diffusion Stochastic Models // J. Royal Stat. Soc. Ser. B (Methodological), Vol. 46, No. 3 (1984), 353-388.

[6] Dynkin, E. B., Markov processes, Vol. 1. Springer, Berlin et al., 1965.

[7] Fakinos, D., The Single-Server Queue with Service Depending on Queue Size and with the Preemptive-Resume Last-Come-First-Served Queue Discipline // Journal of Applied Probability, 1987, Vol. 24, No. 3, 758–767
[8] Gnedenko, B. V., Kovalenko, I. N., Introduction to queueing theory. 2nd ed., rev. and suppl. Birkhäuser, Boston, MA et al., 1991.

[9] Jacobsen, M., Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes, Birkhäuser, Boston, 2006.

[10] Kalashnikov, V. V., Some properties of piecewise linear Markov processes // Theory Probab. Appl., 1975, 20, 560–571.

[11] Krylov, N. V., Introduction to the Theory of Random Processes. AMS, Providence, RI, 1995.

[12] Kulik, A. M., Veretennikov, A. Yu., Diffusion approximation for systems with weakly ergodic Markov perturbations I // Theory Probab. Matem. Stat., 2012. 87, 1–16

[13] Liptser, R. Sh., Shiryaev, A.N., Stochastic calculus on probability spaces with filtration. // Stochastic calculus (S.V Anulova, A.Yu. Veretennikov, N.V.Krylov, R.Sh.Liptser, A.N.Shiryaev), Itogi Nauki i Tekhniki. Modern problems of fundamental math. directions, VINITI, Moscow, 1989, 114–159.

[14] Sevastyanov, B. A., Limiting theorem for Markov processes and its application to telephone systems with losses // Probab. Theory Appl. 1957, 2(1), 106–116.

[15] Thorisson, H., Coupling, stationarity, and regeneration. Springer-Verlag, New York, 2000.

[16] Thorisson, H., The queue GI/G/1: finite moments of the cycle variables and uniform rates of convergence // Stoch. Proc. Appl. 1985, 19(1), 85–99.

[17] Veretennikov, A. Yu., On convergence rate to stationary distribution in Erlang–Sevastyanov systems // Queueing Systems, 2013 (to appear); arXiv.org, 2013, arXiv:1310.5973 [math.PR].

[18] Veretennikov, A. Yu., On convergence rate to stationary distribution in queueing systems with a single server // Avtomatika i telemekhanika, 2013, 74(10), 23-35 (in Russian); (Engl. transl.) Aut. Rem. Control (to appear).

[19] Veretennikov, A. Yu., The ergodicity of service systems with an infinite number of servomechanisms // Math. Notes, 1977, 22(4), 804–808.