SENSITIVITY ERROR BOUNDS FOR NON-EXPONENTIAL STOCHASTIC NETWORKS

Nico M. van Dijk

Stochastic service networks are studied with inaccuracies or perturbations in the distributional forms of service and interarrival times. A condition is provided to conclude error bounds for the effect of these data imprecisions on stationary measures such as throughput. The verification of this condition involves a continuous-state Markov reward recursion relation, which can be performed in an analytic manner. This will be illustrated in detail for a tandem queueing network with imprecisions in a non-exponential input. An explicit error bound on the effect of these imprecisions will be obtained.

INTRODUCTION

Background. Stochastic networks have gained a wide popularity over the last decades in telecommunications, computer performance evaluation and flexible manufacturing. Most notably, explicit product form expressions and related insensitivity properties have been intensively investigated (cf. $[3, 4, 11, 14, 40]$). Generally these expressions rely upon assumptions such as Poissonian arrivals, exponential services, special service disciplines (e. g. processor sharing) and reversible or state independent routings. Such assumptions are typical not met in practice. Simulation, numerical or approximation techniques must then be used. As these techniques can be computationally expensive, robustness or sensitivity results with respect to system input data are of interest.

Motivation. Particularly, the distributional forms of interarrival and service times are the key-factor for both computational and sensitivity results. As these forms are usually obtained by experimental data, inaccuracies are naturally involved. Error bounds on the effect of distributional imprecisions are thus of natural interest. Conversely, for simulation or approximation purposes, robustness results may provide one flexibility in choosing convenient distributional forms, for example Weibull (easily invertible) or Erlang (Markov chain analysis).

Literature. Perturbation or sensitivity analysis has recently enjoyed considerable attention in connection with simulation (cf. $[5, 6, 7, 8, 9, 10, 24, 25]$). This analysis, however, provides sensitivity bounds based on simulated sample path outcomes and does not secure a priori bounds. Moreover, only a number of system parameters and not total distributional forms are studied. Analytic perturbation or related truncation results that provide a priori error bounds for stationary characteristics have also been obtained (e.g. $[12, 13, 15, 16, 17, 18, 29, 35]$). Without exception though, these are all concerned with discrete state Markov chains, while no analogues have been reported for continuous-state Markov chains as the present paper requires. Alternatively, for particular examples, insensitivity bounds have been established by modifying the original system in insensitive product form systems (cf. [28, 30, 31]). This techniques, however, is limited to special systems and does not apply to interarrival times. Moreover, these bounds are merely first quick indicators and do not secure orders of accuracy for small data perturbations.

Result. This paper provides an analytical tool to conclude explicit a priori error bounds for the amount of sensitivity due to imprecisions in distributional data for interarrival and service times. The essential step to this end is the estimation of so-called bias terms of an underlying reward structure. Such estimates have recently been established for various discrete state queueing network applications (cf. [31, 32, 35, 36]). The present application though, requires a continuous-state description. Bias terms estimates in that case do not seem to be available. The major part of this paper therefore is concerned with an illustration of how this can be established analytically in concrete multi-dimensional situations. To this end, a tandem network with non-exponential renewal input is studied. An explicit error bound on the effect of imprecisions or perturbations of this renewal distribution is derived.

1. GENERAL MODEL

For convenience, we first consider the closed case. The open case will be illustrated by the Jacksonian application. Consider a closed stochastic network with a fixed number of M jobs, numbered $1, \ldots, M$. At any moment the state of the system is represented by

 (L, T)

where

$$
\left\{\begin{array}{c} L = (\alpha_1, \ldots, \alpha_M) \\ T = (t_1, \ldots, t_M) \end{array}\right\}
$$

denotes for each job i its current status $\alpha_i \in S$, with S some countable set, and amount of service t_i that it has received since its last service completion.

Example: (Queueing network). In a queueing network we can have: $\alpha_i =$ (r, i, p) denoting for job i its current type number r, the station j at which it is present and a service number p , while t_i is the amount of service that it has already received at this station. Here we note that service numbers have the same function as position numbers but are kept fixed (do not shift) during a service.

Law of motion: The system dynamics are determined by the system character-

as follows: When a job changes its jobmark into ℓ it requires a random amount of service with distribution function \mathbf{F}_{ℓ} . When the system is in state [L, T] the service rate, i.e. the amount of service per unit of time provided to job i is $s_i([L, T])$. When the system is in state $[L, T]$ and job i completes its service, its jobmark is changed into α' with probability $p_i(\alpha' | [L, T]).$

Remarks.

- 1. Note that the service rate for a particular job in a particular state can be equal to zero. This naturally arises for instance in a queueing network with FCFS-service stations, as will be illustrated in the example below.
- 2. Clearly, the above parametrization could have been combined in one service completion rate function. However, the present more detailed formulation is preferred as it corresponds more naturally to queueing network protocols.

Example: (Queueing network). Consider a closed queueing network with N first-come first-served (FCFS)-single server stations and M numbered jobs. A job requires a random amount of service at the various stations, say at station j according to a distribution function G_i . When a job enters a station it is assigned a service number $p + 1$ where p is the largest service number of jobs currently present at that station. This service number remains unchanged until the job completes its service. Under the FCFS-discipline only the job with the smallest service number (which represents the head of the queue) is provided service. The service rate at station j is $b_i(n_i, t_i)$ when n_i jobs are present at this station while the job in service has received already t_j units of service. Upon service completion at station j a job routes to station k with probability $p_{ik}(n, t)$ where $n = (n_1, \ldots, n_N)$ and $t = (t_1, \ldots, t_N)$ just prior to the completion.

Let $\alpha = (i, j, p)$ denote the job-number i of a job, the station number j at which it is present and the service number p of the job at this station. In a given state, let p_i be the smallest service number of jobs present at station j which represents the head of the queue. Also read $1\{A\} = 1$ if even A is satisfied and $1\{A\} = 0$ if not. The above parametrization then applies with

$$
\begin{cases}\n\boldsymbol{F}_{\alpha} = \boldsymbol{G}_{j} \\
\boldsymbol{s}_{i}([L, T]) = \boldsymbol{b}_{j}(n_{j}, t_{j}) \boldsymbol{1}(p = p_{j}) \\
\boldsymbol{p}_{i}(\alpha' | [L, T]) = \boldsymbol{p}_{jk}(\boldsymbol{n}, t) \boldsymbol{1}(\alpha' = (i, k, p)).\n\end{cases}
$$

The following assumptions are made in order to define a convenient transformation.

Assumptions:

1. For all α , the function $\mathbf{F}_{\alpha}(t)$ is absolutely continuous for $t \in (0,\infty)$ with density function $f_{\alpha}(t)$. Hence, its failure rate, say $h_{\alpha}(t)$, is well-defined by $h_{\alpha}(t) = \mathbf{f}_{\alpha}(t)/[1 - \mathbf{F}_{\alpha}(t)]$ for all $t \in (0, \infty)$. We introduce the notation:

$$
\boldsymbol{d}_{i}([L,T]) = \boldsymbol{s}_{i}([L,T]) \, \boldsymbol{f}_{\alpha_{i}}(t_{i}) / [1 - \boldsymbol{F}_{\alpha_{i}}(t_{i})] \tag{1}
$$

2. For some constant $B < \infty$ and all $[L, T]$:

$$
d([L, T]) = \sum_{i} d_i([L, T]) \le B.
$$
 (2)

Uniformized model: Let Q be an arbitrary finite number with $Q > B$. We now define a related continuous-time Markov chain model as follows. At exponential times with parameter Q the system will have a jump. For $v > 0$, let $T + v$ denote the vector $(t_1 + v, t_2 + v, \ldots, t_M + v)$. If directly after the last jump the system was in state $[L, T]$ while the next jump will take place after time v, by this next jump with probability

$$
\boldsymbol{P}_{v}([L,T], [L',T']) = \boldsymbol{d}_{i}([L,T+v]) \, \boldsymbol{p}_{i}(\alpha' | [L,T+v]) \, / \, Q \tag{3}
$$

the system will change into state $[L', T']$ with $\alpha'_j = \alpha_j$ and $t_j = t_j + v$ for $j \neq i$ but $\alpha'_j = \alpha'$ and $t'_i = 0$, for all $i = 1, \ldots, M$. With probability

$$
1 - \mathbf{d}([L, T + v]) / Q \tag{4}
$$

only the ages are updated, i.e. the state will change into $[L, T + v]$. Without loss of generality, assume that both the original and the above uniformized model have a unique stationary density at one and the same irreducible set of states R which we denote by $\pi_1(L,T)$ and $\pi_2(L,T)$ respectively. The following lemma is proven in [33] and will enable us to restrict to a discrete step Markov chain analysis such as to employ inductive arguments.

Lemma 1.1. $\pi_1(\cdot) = \pi_2(\cdot).$

2. SENSITIVITY BOUNDS

Consider a similar perturbed stochastic network with the characteristics

$$
\mathbf{F}_{\alpha}(\cdot), \ \mathbf{s}_{i}(\cdot), \ \mathbf{p}_{i}(\cdot | \cdot)
$$
 replaced by
$$
\overline{\mathbf{F}}_{\alpha}(\cdot), \ \overline{\mathbf{s}}_{i}(\cdot), \ \overline{\mathbf{p}}_{i}(\cdot | \cdot)
$$

and without loss of generality assume that (1) and (2) hold again with the same value Q. Then by virtue of Lemma 1.1, its stationary distribution, denoted by $\overline{\pi}(\cdot)$ is also determined by the uniformized model with the above substitutions. From now on, we will always denote an expression for this perturbed system with an upper bar symbol '-'. Further, we only give definitions for the original system while those for the perturbed system are analogues.

Let $\mathbf{R}(L,T)$ be some reward function and define functions $\overline{\mathbf{V}}_n(\cdot)$ and $\mathbf{V}_n(\cdot)$ for $n = 1, 2, \dots$ by: $\overline{V}_0 = V_0(\cdot) = 0$ and for $n = 1, 2, \dots$:

$$
\boldsymbol{V}_{n+1}(L,T) = \boldsymbol{R}(L,T) + \int_0^\infty Q \, e^{-vQ} \sum_{[L',T']} \boldsymbol{P}_v([L,T],[L',T']) \, \boldsymbol{V}_n(L',T') \, dv, \tag{5}
$$

where it is to be noted that the summation over $[L',T']$ for fixed value v actually comes down to summation over all possible components i which determines which component will change or, $[L', T'] = [L, T + v]$. In words that is, $V_n(L, T)$ represents the total expected rewards over n exponential periods with parameter Q under the one-step transition structure $P_v(\cdot, \cdot)$ and one-step rewards $R(\cdot, \cdot)$ and given the initial state $[L, T]$ at time 0. Now assume that for some initial state $[L_0, T_0]$

$$
G = \lim_{N \to \infty} \frac{Q}{N} \boldsymbol{V}_N(L_0, T_0)
$$
\n⁽⁶⁾

exists and is well-defined. As $R(\cdot, \cdot)$ represents a one-step reward per period of expected length Q^{-1} , the values \overline{G} then represent an expected reward per unit of time when the system is in equilibrium. For example, for some given reward rate function $r(L, T)$ we can have

$$
\boldsymbol{R}(L,T) = \int_0^\infty Q \, e^{-vQ} \boldsymbol{r}(L,T+v) \, \mathrm{d}v \tag{7}
$$

as corresponding to a reward measurement just prior to jumps, or

$$
\mathbf{R}(L,T) = \int_0^\infty Q e^{-vQ} \left[\int_0^v \mathbf{r}(L,T+s) \,ds \right] dv
$$

as a reward rate measurement continuously in time. The following key-theorem can now be formulated. Herein, let $[\overline{P}_v - P_v]$ $([\cdot, \cdot], [\cdot, \cdot']) = \overline{P}_v([\cdot, \cdot], [\cdot, \cdot'])$ $\boldsymbol{P}_{v}([\cdot,\cdot],[\cdot,\cdot'])$.

Theorem 2.1. Suppose that for some Δ_1 , $\Delta_2 \ge 0$ and all [L, T], and $n \ge 0$:

$$
\left| \sum_{[L',T']} \left[\overline{\boldsymbol{P}}_v - \boldsymbol{P}_v \right] \left([L,T], \, [L',T'] \right) \left[\boldsymbol{V}_n(L',T') - \boldsymbol{V}_n(L,T) \right] \right| \leq \Delta_1/Q \qquad (8)
$$

$$
\left|\overline{R}(L,T) - R(L,T)\right| \leq \Delta_2/Q. \tag{9}
$$

Then

$$
|\overline{G} - G| \le \Delta_1 + \Delta_2. \tag{10}
$$

P r o o f. By virtue of (5) :

$$
\left(\overline{\mathbf{V}}_{n+1} - \mathbf{V}_{n+1}\right) ([L, T]) = \left[\overline{\mathbf{R}}(L, T) - \mathbf{R}(L, T)\right]
$$
\n
$$
+ \int_0^\infty Q \, e^{-vQ} \Bigg\{ \sum_{[L', T']} \left[\overline{\mathbf{P}}_v([L, T], [L', T']) - \mathbf{P}_v([L, T], [L', T')] \mathbf{V}_n(L', T')\right. \\ \left. + \sum_{[L', T']} \overline{\mathbf{P}}_v([L, T], [L', T']) \left[\overline{\mathbf{V}}_n(L', T') - \mathbf{V}_n(L', T')\right] \, dv.
$$
\n
$$
(11)
$$

Noting that

$$
\sum_{[L',T']} \overline{P}_v([L,T],[L',T']) = 1,
$$

we have

$$
\sum_{[L',T']} \left[\overline{P}_v([L,T],[L',T']) - P_v([L,T],[L',T']) \right] \mathbf{V}_n(L',T')
$$
\n
$$
= \sum_{[L',T']} \left[\overline{P}_v([L,T],[L',T'] - P_v([L,T],[L',T'])] \right] \tag{12}
$$
\n
$$
\left[\mathbf{V}_n(L',T') - \mathbf{V}_n(L,T) \right].
$$

Substituting (12) in (11) and applying (8) and (9) yields for any $[L, T]$:

$$
\gamma_{n+1} = \sup_{[L,T]} |\overline{V}_{n+1}(L,T) - V_{n+1}(L,T)|
$$

$$
\leq \int_0^\infty \Delta_1 e^{-vQ} dv + \Delta_2 Q^{-1} + \sup_{[L',T']} [\overline{V}_n(L',T') - V_n(L',T')] \leq \gamma_n + [\Delta_1 + \Delta_2] Q^{-1}.
$$

Iterating this expression for $n = N - 1, ..., 0$ and substituting $\overline{V}_0(\cdot) = V_0(\cdot) = 0$ gives for any $[L, T]$: \overline{a}

$$
\left|\overline{\boldsymbol{V}}_{N}(L,T)-\boldsymbol{V}_{N}(L,T)\right|\leq\left[N/Q\right]\left[\boldsymbol{\Delta}_{1}+\boldsymbol{\Delta}_{2}\right].
$$

Inserting $[L, T] = [L_0, T_0]$ and applying (6) completes the proof. \Box

Corollary 2.1. Let
$$
\overline{R}(\cdot, \cdot) = R(\cdot, \cdot)
$$
\n
$$
\overline{s}_i(\cdot) = s_i(\cdot) \leq S
$$
\n
$$
\overline{p}_i(\cdot | \cdot) = p_i(\cdot | \cdot)
$$
\n
$$
\overline{h}_{\alpha}(t) = \overline{f}_{\alpha}(\cdot) / [1 - \overline{F}_{\alpha}(t)]
$$
\n(13)

and assume that for some constant δ and $\boldsymbol{C}\geq 0:$

$$
\left| \overline{h}_{\alpha}(t) - h_{\alpha}(t) \right| \le \delta \tag{14}
$$

$$
|\boldsymbol{V}_n(L',T') - \boldsymbol{V}_n(L,T)| \leq \boldsymbol{C}
$$
\n(15)

for all α and $n, t \geq 0$ and $[L',T']$ with $(\alpha'_j,t'_j) = (\alpha_j,t_j)$ for $j \neq i$ while $t'_i = 0$ for some $i \in \{1, \ldots, M\}$. Then

$$
|\overline{G} - G| \le \delta \, C \, S. \tag{16}
$$

Remark 2.1 (Monotonicity results). The proof can almost be reread identically to conclude monotonicity results of the form

$$
\overline{G}\leq G\quad\text{or}\quad \overline{G}\geq G
$$

when (8) holds without absolute values and the right hand side replaced by ≤ 0 or ≥ 0 . Monotonicity results for queueing networks have been extensively studied over the last decade such as with respect to the number of servers or jobs (e. g. $[1, 2, 19, 20, 21, 22, 23, 26, 28, 36, 39]$. Monotonicity results with respect to distributional forms though are limited to some results for simple Erlang type facilities (cf. [22, 38]). As such, the above results in monotonicity form would be an extended form of possible interest. The primary focus herein, however, are error bounds.

3. APPLICATION: A TANDEM NETWORK WITH NON-EXPONENTIAL INPUT

To illustrate how the necessary condition (8) or (15) can be verified in concrete situations, this section investigates a particular application: A finite tandem line. As a non-exponential input is a realistic phenomenon but also known to be a key-factor of the failure for an explicit product form results, we particularize this application to a non-exponential input while for convenience of presentation service times are assumed to be exponential.

The system under study is an open two station tandem line with a finite capacity constraint of no more than N jobs. Jobs arrive at the system according to a renewal input with interarrival (renewal) distribution $\mathbf{F}(\cdot)$. When the system is saturated, i.e. $n = N$ where n is the number of jobs already present, an arriving job is rejected and lost. Otherwise it enters station 1. After service completion at station 1, a job instantly routes to station 2 and after service completion at station 2 it directly leaves the system. When n_i jobs are present at station i the rate at which jobs are completed is $\mu_i(n_i)$, where $\mu_i(n_i)$ is assumed to be non-decreasing in n_i , $i = 1, 2$, where we assume services to be exponential.

The state of the system can be described by $[n, t]$, where the vector $n = (n_1, n_2)$ denotes the numbers n_i of jobs at stations $i = 1, 2$ and where t is the time after the last arrival. The results of Section 2 do not apply directly as the number of jobs is not fixed and external arrivals are involved. A way to include open models in the description of Section 2 is to let arriving jobs be assigned an arrival number, to be included in the status α of a job, and to use a special number 0 to describe an external job which can enter the system of which is created when a job leaves the system. However, for the special system under consideration, we prefer to give a somewhat more direct version. More precisely, consider the system described above but with the interarrival distribution modified in $\overline{F}(\cdot)$ and let

$$
\begin{cases} \boldsymbol{h}(t) = \boldsymbol{f}(t) / [1 - \boldsymbol{F}(t)] \\ \overline{\boldsymbol{h}}(t) = \overline{\boldsymbol{f}}(t) / [1 - \overline{\boldsymbol{F}}(t)] \end{cases}
$$

be the corresponding arrival failure rates for the original and modified system, which are assumed to be well-defined for all $t \in (0, \infty, \text{ Now}, \text{ with})$

$$
\mathbf{r}(\mathbf{n}, t+v) = \mathbf{h}(t+v)
$$

and

$$
Q \ge \sup_t 2h(t) + \sup_{\mathbf{n}} [\mu_1(n_1) + \mu_2(n_2)]
$$
 (17)

for $z = 0, 1, 2, \ldots$ define functions $\boldsymbol{V}_z(\boldsymbol{n}, t)$ as per (5) and (7). More precisely, define $\boldsymbol{V}_0(\cdot) = 0$ and

$$
\mathbf{V}_{m+1}(\boldsymbol{n},t) = \int_0^\infty e^{-vQ} \left\{ \boldsymbol{h}(t+v) \mathbf{1}_{\{n\n(18)
$$

The value G as defined by

$$
G = \lim_{z \to \infty} \frac{Q}{z} V_z(0,0), \tag{19}
$$

where $\mathbf{0} = (0, 0)$, then represents the total system throughput, that is the mean number of accepted jobs or system departures per unit of time, when the system is in equilibrium. The functions $\bar{r}(n, t + v)$ and $\bar{V}_z(n, t)$ and the value \bar{G} are defined similarly for the perturbed model. Now similarly to Theorem 2.1 and using the fact that the systems differ in only their arrival failure rates, we can prove

Result 3.1.

$$
|\boldsymbol{G} - \overline{\boldsymbol{G}}| \le \delta[1 + \boldsymbol{C}] \tag{20}
$$

when for all $n + e_1$ and $m, t \geq 0$:

$$
|\mathbf{h}(t) - \overline{\mathbf{h}}(t)| \le \delta,\tag{21}
$$

$$
|\boldsymbol{V}_m(\boldsymbol{n}+e_1, 0) - \boldsymbol{V}_m(\boldsymbol{n}, t)| \le C. \tag{22}
$$

As condition (21) is determined by the system data or modelling, the essential condition to be verified is (22). The following result proves the concrete simple estimate $C = 1$ when the arrival failure rate is monotone non-decreasing.

Result 3.2. Assume that $h(t)$ is non-decreasing in t. Then for all n, t, s, i and z:

$$
0 \leq \delta_s \mathbf{V}_s(\mathbf{n}, t) = \mathbf{V}_z(\mathbf{n}, t + s) - \mathbf{V}_z(\mathbf{n}, t) \leq 1, \tag{23}
$$

 $0 \ge \Delta_1^s \, \boldsymbol{V}_z(\boldsymbol{n},t) = \boldsymbol{V}_z(\boldsymbol{n} + e_1, t) - \boldsymbol{V}_z(\boldsymbol{n}, t + s) \ge -1,$ (24)

$$
0 \ge \Delta_2^s \, \mathbf{V}_z(\mathbf{n}, t) = \mathbf{V}_z(\mathbf{n} + e_2, t) - \mathbf{V}_z(\mathbf{n}, t + s) \ge -1,\tag{25}
$$

$$
0 \ge \Delta_3^s \, \mathbf{V}_z(\mathbf{n}, t) = \mathbf{V}_z(\mathbf{n} + e_1, t) - \mathbf{V}_z(\mathbf{n} + e_2, t + s) \ge -1. \tag{26}
$$

P r o o f. This will be given by induction to z. Clearly (23) – (26) hold for $z = 0$ as $V_0(\cdot) = 0$. Suppose that $(23) - (26)$ hold for $z = m$. Below we will then express V_{m+1} in V_m by means of (18). Before doing so it is noted in advance that in the derivations that follow, some terms are artificially added and subtracted (e. g. $h(t + s + v) - h(t + v)$ in (27)) and $\mu_1(n_1 + 1) - \mu_1(n_1)$ in (28)) or artificially split (e. g. $h(t+s+v)$ in $h(t+v) + [h(t+s+v) - h(t+v)]$ in (27) and $\mu_i(n_i+1) =$ $\mu_i(n_i)+[\mu_i(n_1+1)-\mu_i(n_i)]$ in (28) in order to compare corresponding terms pairwise with equal coefficients. Further, as the detailed technicalities are slightly different but crucial, the derivations will be given in full detail for all inequalities to be proven.

$$
\delta_{s} \mathbf{V}_{m+1}(\mathbf{n},t) \qquad (27)
$$
\n
$$
= \int_{0}^{\infty} e^{-vQ} \left\{ \mathbf{h}(t+s+v) \mathbf{1}_{\{n\n
$$
- \int_{0}^{\infty} e^{-vQ} \left\{ \mathbf{h}(t+v) \mathbf{1}_{\{n\n
$$
- \int_{0}^{\infty} e^{-vQ} \left\{ \left[\mathbf{h}(t+s+v) - \mathbf{h}(t+v) \right] \mathbf{1}_{\{n
$$
$$
$$

Now note that by induction hypothesis (24) for $z = m$, the second term between braces $\{\cdot\}$ in the latter expression can be negative but is bounded from below by $-[h(t+s+v)-h(t+v)] 1_{\{n\leq N\}}$. By combining this negative estimate with the first positive term, which is exactly the same up to sign, applying the induction hypothesis $\delta_s V_m(\cdot) \geq 0$ and recalling (23), one concludes: $\delta_s V_m(\cdot) \geq 0$. To estimate the latter expression from above by 1, delete the second term which is non-positive by induction assumption, apply the induction hypothesis $\delta_s V_m(\cdot) \leq 1$ and note that all terms between braces $\{\cdot\}$ now sum up to 1 by virtue of (17). Inequality (23) is hereby verified for $z = m + 1$.

To verify (24), again we will apply (18) where the remarks made above are re-

called. We then obtain

$$
\begin{split}\n&\Delta_{1}^{s}V_{m+1}(n,t) \\
&= \int_{0}^{\infty} e^{-vQ} \left\{ h(t+v) \mathbf{1}_{\{n+1
$$

Here note all $\delta_{[t+v]}$ $V_m(\cdot)$ are non-negative by induction hypothesis (23) for $z = m$, but estimated from above by 1. As a consequence, by substituting these upper estimates, combining them with the first two negative terms, which are exactly equal to their coefficients up to sign, applying the induction hypothesis $\Delta^s_2 \, \boldsymbol{V}_m(\cdot) \leq 0$ and $\Delta_1^s V_m(\cdot) \leq 0$ and recalling (17), we conclude: $\Delta_1^s V_{m+1}(\cdot) \leq 0$. To estimate the latter expression from below by -1 , delete all $\delta_{[t+v]} V_m(\cdot)$ -terms, which are nonnegative by induction assumption, apply the induction hypotheses $\Delta_2^s V_m(\cdot) \geq -1$ and $\Delta_1^s V_m(\cdot) \geq -1$ and note that all terms between braces $\{\cdot\}$ now sum up to -1 by virtue of (17). Inequality (24) is hereby verified for $z = m + 1$. To prove (25), we obtain similarly to (27):

$$
\Delta_{2}^{s}V_{m+1}(n,t) \qquad (29)
$$
\n
$$
= \int_{0}^{\infty} e^{-vQ} \left\{ h(t+v) 1_{\{n+1\n
$$
= \int_{0}^{\infty} e^{-vQ} \left\{ h(t+s+v)] 1_{\{n+1
$$
$$

Now note that all $-\Delta_3 V_m(\cdot)$ -terms are non-negative as per induction hypothesis (26) for $z = m$ but estimated from above by 1. Hence, as in (27) by substituting these upper estimates, combining them with the first two negative terms which are exactly equal to their coefficients up to sign, applying the induction hypotheses $-\delta_s \mathbf{V}_m(\cdot) \leq 0$ and $\Delta_2^s \mathbf{V}_m(\cdot) \leq 0$ and recalling (17) we conclude: $\Delta_2^s \mathbf{V}_{m+1}(\cdot) \leq 0$.

Conversely, as before, by deleting the non-negative $-\Delta_3^s V_m(\cdot)$ terms, applying $-\delta_s V_m(\cdot) \geq -1$ and $\Delta_2^s V_m(\cdot) \geq -1$ as per hypotheses and noting that all terms between braces then sum up to -1 by virtue of (17) we obtain $\Delta_2^s V_{m+1}(\cdot) \geq -1$. Inequality (25) is thus proven for $z = m + 1$. Finally, again as in (28) we conclude:

$$
\Delta_{3}^{s} V_{m}(n,t) = V_{m+1}(n+e_{1}, t) - V_{m+1}(n+e_{2}, t+s)
$$
\n
$$
= \int_{0}^{\infty} e^{-vQ} \left\{ h(t+v) \mathbf{1}_{\{n+1\n
$$
+ h(t+v) \mathbf{1}_{\{n+1\n
$$
+ [h(t+s+v) - h(t+v)] \mathbf{1}_{\{n+1\n
$$
+ [\mu_{1}(n_{1}+1) - \mu_{1}(n_{1})] V_{m}(n+e_{1}, t+v) + \mu_{1}(n_{1}) V_{m}(n+e_{2}, t+v)
$$
\n
$$
+ [\mu_{2}(n_{2}+1) - \mu_{2}(n_{2})] V_{m}(n+e_{1}, t+v) + \mu_{2}(n_{2}) V_{m}(n+e_{1}-e_{2}, t+v)
$$
\n
$$
+ [Q-h(t+s+v) \mathbf{1}_{\{n+1\n
$$
- \int_{0}^{\infty} e^{-vQ} \left\{ h(t+s+v) \mathbf{1}_{\{n+1\n
$$
+ [h(t+s+v) - h(t+v)] \mathbf{1}_{\{n+1\n
$$
+ [\mu_{1}(n_{1}+1) - \mu_{1}(n_{1})] V_{m}(n+e_{2}, t+s+v) + \mu_{1}(n_{1}) V_{m}(n-e_{1}+e_{2}+e_{2}, t+s+v)
$$
\n
$$
+ [\mu_{2}(n_{2}+1) - \mu_{2}(n_{2})] V_{m}(n, t+s+v) + \mu_{2}(n_{2}) V_{m}(n, t+s+v)
$$
\n
$$
+ [\mu_{2}(n_{2}+1) - \mu_{2}(n_{2})] V_{m}(n+e_{1}, t+s+v) + \mu_{2}(n_{2}) V_{m}(n, t+s+v)
$$
\
$$
$$
$$
$$
$$
$$

Here the $-\Delta_2 V_m(\cdot)$ term is non-negative but estimated from above by 1 as per induction hypothesis (25) for $z = m$. Hence, as before, by substituting this upper estimate, combining it with the first negative term which is exactly equal to its coefficient up to sign, applying the hypotheses: $\Delta_3^s V_m(\cdot) \leq 0$, and $-\delta_s V_m(\cdot) \leq 0$ and recalling (17) we conclude: $\Delta_3^s V_{m+1}(\cdot) \leq 0$.

Conversely, by deleting the non-negative $-\Delta_2^s\, \boldsymbol{V}_m(\cdot)$ term, applying $-\delta_s\, \boldsymbol{V}_m(\cdot) \geq 0$ −1 and $\Delta_3^s V_m(\cdot) \ge -1$ as per hypotheses and noting that all terms between braces { \cdot } then sum up to −1 by virtue of (17), we obtain $\Delta_3 V_{m+1}(\cdot) \ge -1$. Inequality (26) is thus proven for $z = m + 1$. By induction the proof of the lemma is hereby completed. $\hfill \square$

By Results 3.1 and 3.2 we thus conclude:

Corollary 3.3. Assuming that $h(t)$ is nondecreasing in t we have under (21):

$$
|\mathbf{G} - \overline{\mathbf{G}}| \le 2\delta. \tag{31}
$$

Remarks (Nondecreasing $h(t)$).

- (i) Note that only $h(t)$ and not $\overline{h}(t)$ is required to be nondecreasing for Corollary 3.3. For example, with $h(t) = \mu$ we can so investigate the effect of a small deviation, as modeled by $h(t)$, from an exponential input assumption.
- (ii) The assumption of a nondecreasing failure rate $h(t)$ is quite realistic. For instance, one can think of an arrival as representing a broken down component where the rate of a component to go down increases by its lifetime.
- (iii) Extensions of Result 3.2 such that $h(t)$ is not necessarily nondecreasing do seem possible along the same lines, but will be technically more complex. Particularly, a weighted mixture of decreasing and nondecreasing failure rates does seem possible.

CONCLUSION

The sensitivity of system performance measures with respect to its underlying stochastic assumptions is an important practical aspect as detailed data are often not available. This paper provides a method by which the effect of impressions or perturbations in underlying distributional forms for arrivals and services in a stochastic service network can be evaluated in analytical manner. This method requires a technical verification for bounding so-called bias terms. To this end, an inductive prooftechnique can be employed based on Markov reward structures. This technique has so far only been applied for discrete-state description. In the present paper it also appears executable for continuous state descriptions, as was illustrated for a tandem queueing system with imprecisions in the arrival distribution.

Further exploration of this method is suggested for more complex networks, for imprecisions in several distributions simultaneously and for situations with wildly varying distributions.

(Received November 8, 1993.)

REFERENCES

- [1] I. J. B. F. Adan and J. van der Wal: Monotonicity of the throughput of a closed queueing network in the number of jobs. Oper. Res. 37 (1989), 953–957.
- [2] I. J. B. F. Adan and J. van der Wal: Monotonicity of the throughput in single server production and assembly networks with respect to the buffer sizes. In: Proceedings of the 1st International Workshop on Queueing Systems with Blocking. North–Holland 1989, pp. 345–356.
- [3] A. D. Barboúr: Networks of queues and the method of stages. Adv. in Appl. Probab. 8 (1976), 584–591.
- [4] J.W. Cohen: Sensitivity and insensitivity. Delft Progr. Rep. 5 (1980), 159–173.
- [5] X. R. Cao: Convergence of parameter sensitivity estimates in a stochastic experiment. IEEE Trans. Automat. Control 30 (1985), 834–843.
- [6] X. R. Cao: First-order perturbation analysis of a single multi-class finite source queue. Performance Evaluation 7 (1987), 31–41.
- [7] X. R. Cao and Y. C. Ho: Sensitivity estimate and optimization of throughput in a production line with blocking. IEEE Trans. Automat. Control 32 (1987), 959–967.
- [8] Y. C. Ho and C. Cassandras: Infinitesimal and finite perturbation analysis for queueing networks. Automatica 19 (1983), 4, 439–445.
- [9] Y. C. Ho and S. Li: Extensions of infinitesimal perturbation analysis. IEEE Trans. Automat. Control 33 (1988), 427–438.
- [10] P. Glasserman and Y. C. Ho: Aggregation approximations for sensitivity analysis of multi-class queueing networks. Performance Evaluation.
- [11] A. Hordijk and N. M. van Dijk: Adjoint process, job-local balance and intensitivity of stochastic networks. Bull. 44th Session Int. Inst. 50 (1983), 776–788.
- [12] C. D. Meyer, Jr.: The condition of a finite Markov chain and perturbation bounds for the limiting probabilities. SIAM J. Algebraic Discrete Methods 1 (1980), 273–283.
- [13] J. R. Rohlicek and A. S. Willsky: The reduction of perturbed Markov generators: an algorithm exposing the role of transient states. J. Assoc. Comput. Mach. 35 (1988), 675–696.
- [14] R. Schassberger: The intensitity of stationary probabilities in networks of queues. Adv. in Appl. Probab. 10 (1987), 906–912.
- [15] P. J. Schweitzer: Perturbation theory and finite Markov chains. J. Appl. Probab. 5 (1968), 401–413.
- [16] E. Seneta: Finite approximations to finite non-negative matrices. Cambridge Stud. Philos. 63 (1967), 983–992.
- [17] E. Seneta: The principles of truncations in applied probability. Comment. Math. Univ. Carolin. 9 (1968), 533–539.
- [18] E. Seneta: Non-Negative Matrices and Markov Chains. Springer Verlag, New York 1980.
- [19] J. G. Shantikumar and D. D. Yao: Stochastic monotonicity of the queue lengths in closed queueing networks. Oper. Res. 35 (1987), 583–588.
- [20] J. G. Shantikumar and D. D. Yao: Throughput bounds for closed queueing networks with queue-dependent service rates. Performance Evaluation $9(1987)$, 69–78.
- [21] J. G. Shantikumar and D. D. Yao: Monotonicity properties in cyclic queueing networks with finite buffers. In: Proceedings of the First Int. Workshop on Queueing Networks with Blocking. North Carolina 1988.
- [22] D. Stoyan: Comparison Methods for Queues and Other Stochastic Models. J. Wiley, New York 1983.
- [23] R. Suri: A concept of monotonicity and its characterization for closed queueing networks. Oper. Res. 33 (1985), 606–024.
- [24] R. Suri: Infinitesimal perturbation analysis for general discrete event systems. J. Assoc. Comput. Mach. 34 (1987), 3, 686–717.
- [25] R. Suri: Perturbation analysis. The state of the art and research issues explained via the GI/GI/1 queue. In: Proceedings of the IEEE.
- [26] H. C. Tijms: Stochastic Modelling and Analysis. A Computational Approach. J. Wiley, New York 1986.
- [27] P. Tsoucas and J. Walrand: Monotonicity of throughput in non-markovian networks. J. Appl. Probab. (1983).
- [28] N. M. van Dijk: A formal proof for the insensivity of simple bounds for finite multiserver non-exponential tandem queues based on monotonicity results. Stochastic Process. Appl. 27 (1988), 216–277.
- [29] N. M. van Dijk: Perturbation theory for unbounded Markov reward process with applications to queueing. Adv. in Appl. Probab. 20 (1988), 99–111.
- [30] N. M. van Dijk: Simple performance bounds for non-product form queueing networks. In: Proceedings of the First International Workshop on Queueing Networks with Blocking. North–Holland 1988, pp. 1–18.
- [31] N. M. van Dijk: Simple bounds for queueing systems with breakdowns. Performance Evaluation 7 (1988), 117–128.
- [32] N. M. van Dijk: A simple throughput bound for large closed queueing networks with finite capacities. Performance Evaluation 10 (1988), 153-167.
- [33] N. M. van Dijk: A note on extended uniformization for non-exponential stochastic networks. J. Appl. Probab. 28 (1991), 955–961.
- [34] N. M. van Dijk and B. F. Lamond: Bounds for the call congestion of finite single-server exponential tandem queues. Oper. Res. 36 (1988), 470–477.
- [35] N. M. van Dijk and M. L. Puterman: Perturbation theory for Markov reward processes with applications to queueing systems. Adv. in Appl. Probab. 20 (1988), 79–99.
- [36] N. M. van Dijk and J. van der Wal: Simple bounds and monotonicity results for multiserver exponential tandem queues. Queueing Systems Theory Appl. 4 (1989), 1–16.
- [37] E.W. B. van Marion: Influence of holding time distributions or blocking probabilities of a grading. TELE 20 (1968), 17–20.
- [38] W. Whitt: Comparing counting processes and queues. Adv. in Appl. Probab. 13 (1981), 207–220.
- [39] W. Whitt: Stochastic comparison for non-Markov processes. Math. Oper. Res. 11 (1986), 4, 608–618.
- [40] P. Whittle: Partial balance and insensitivity. Adv. in Appl. Probab. 22 (1985), 168– 175.

Prof. Dr. Nico M. van Dijk, Faculty of Economical Sciences and Econometrics, University of Amsterdam, Roetersstraat 11, 1 018 WB Amsterdam. The Netherlands.